Chapter 30

The RELIABILITY Procedure

Chapter Table of Contents

Chapter 30
The RELIABILITY Procedure

Overview

The RELIABILITY procedure provides tools for reliability and survival data analysis and for recurrence data analysis. You can use this procedure to

- construct probability plots and fitted life distributions with left-, right-, and interval-censored lifetime data
- fit regression models, including accelerated life test models, to combinations of left-, right-, and interval-censored data
- analyze recurrence data from repairable systems

These tools benefit reliability engineers and industrial statisticians working with product life data and system repair data. They also aid workers in other fields, such as medical research, pharmaceuticals, social sciences, and business, where survival and recurrence data are analyzed.

Most practical problems in reliability data analysis involve right-censored or interval-censored data. The RELIABILITY procedure provides probability plots of uncensored, right-censored, and interval-censored data when all the failure data have common interval endpoints.

Features of the RELIABILITY procedure include

- probability plotting and parameter estimation for the common life distributions: Weibull, exponential, extreme value, normal, lognormal, logistic, and loglogistic. The data can be complete, right censored, or interval censored.
- maximum likelihood estimates of distribution parameters, percentiles, and reliability functions
- both asymptotic normal and likelihood ratio confidence intervals for distribution parameters and percentiles. Asymptotic normal confidence intervals for the reliability function are also available.
- estimation of distribution parameters by least squares fitting to the probability plot
- Weibayes analysis, where there are no failures and where the data analyst specifies a value for the Weibull shape parameter
- estimates of the resulting distribution when specified failure modes are eliminated

- plots of the data and the fitted relation for life versus stress in the analysis of accelerated life test data
- fitting of regression models to life data, where the life distribution location parameter is a linear function of covariates. The fitting yields maximum likelihood estimates of parameters of a regression model with a Weibull, exponential, extreme value, normal, lognormal, logistic and loglogistic, or generalized gamma distribution. The data can be complete, right censored, left censored, or interval censored. For example, accelerated life test data can be modeled with such a regression model.
- nonparametric estimates and plots of the mean cumulative function for cost or number of repairs and associated confidence intervals from repair data from systems

Some of the features provided in the RELIABILITY procedure are available in other SAS procedures.

- You can construct probability plots of life data with the CAPABILITY procedure; however, the CAPABILITY procedure is intended for process capability analysis rather than reliability analysis, and the data must be complete, that is, uncensored.
- The LIFEREG procedure fits regression models with life distributions such as the Weibull, lognormal, and loglogistic to left-, right-, and interval-censored data. The RELIABILITY procedure fits the same distributions and regression models as the LIFEREG procedure and, in addition, provides a graphical display of life data in probability plots.

Lawless (1982), Nelson (1990), Nelson (1982), and Tobias and Trindade (1995) provide many examples taken from diverse fields and describe the analyses provided by the RELIABILITY procedure. Nelson emphasizes reliability data analysis from an engineering viewpoint.

The features of the procedure that deal with the analysis of repair data from systems are based on the work of Nelson (1995), Nelson (1988), Doganaksoy and Nelson (1991), and Nelson and Doganaksoy (1989), who provide examples of repair data analysis.

Getting Started

This section introduces the RELIABILITY procedure with examples that illustrate some of the analyses that it performs.

Analysis of Right-Censored Data from a Single Population

The Weibull distribution is used in a wide variety of reliability analysis applications. This example illustrates the use of the Weibull distribution to model product life data from a single population. The following statements create a SAS data set containing observed and right-censored lifetimes of 70 diesel engine fans (Nelson 1982, p. 318).

```
data fan;
   input lifetime censor@@;
   lifetime = lifetime / 1000;
   datalines;
    450 0    460 1   1150 0   1150 0   1560 1
   1600 0   1660 1   1850 1   1850 1   1850 1
   1850 1   1850 1   2030 1   2030 1   2030 1
   2070 0   2070 0   2080 0   2200 1   3000 1
   3000 1   3000 1   3000 1   3100 0   3200 1
   3450 0   3750 1   3750 1   4150 1   4150 1
   4150 1   4150 1   4300 1   4300 1   4300 1
   4300 1   4600 0   4850 1   4850 1   4850 1
   4850 1   5000 1   5000 1   5000 1   6100 1
   6100 0   6100 1   6100 1   6300 1   6450 1
   6450 1   6700 1   7450 1   7800 1   7800 1
   8100 1   8100 1   8200 1   8500 1   8500 1
   8500 1   8750 1   8750 0   8750 1   9400 1
   9900 1  10100 1  10100 1  10100 1  11500 1
   ;
run;
```

Some of the fans had not failed at the time the data were collected, and the unfailed units have right-censored lifetimes. The variable LIFETIME represents either a failure time or a censoring time in thousands of hours. The variable CENSOR is equal to 0 if the value of LIFETIME is a failure time, and it is equal to 1 if the value is a censoring time.

The following statements use the RELIABILITY procedure to produce the graphical output shown in Figure 30.1:

```
proc reliability;
   distribution weibull;
   probplot lifetime*censor( 1 ) / covb;
run;
```

The DISTRIBUTION statement specifies the Weibull distribution for probability plotting and maximum likelihood (ML) parameter estimation. The PROBPLOT statement produces a probability plot for the variable LIFETIME and specifies that the value of 1 for the variable CENSOR denotes censored observations. You can specify any value, or group of values, for the *censor-variable* (in this case, CENSOR) to indicate censoring times. The option COVB requests the ML parameter estimate covariance matrix.

The graphical output, displayed in Figure 30.1, consists of a probability plot of the data, an ML fitted distribution line, and confidence intervals for the percentile (lifetime) values. An *inset* box containing summary statistics, Weibull scale and shape estimates, and other information is displayed on the plot by default. The locations of the right-censored data values are plotted in an area at the top of the plot.

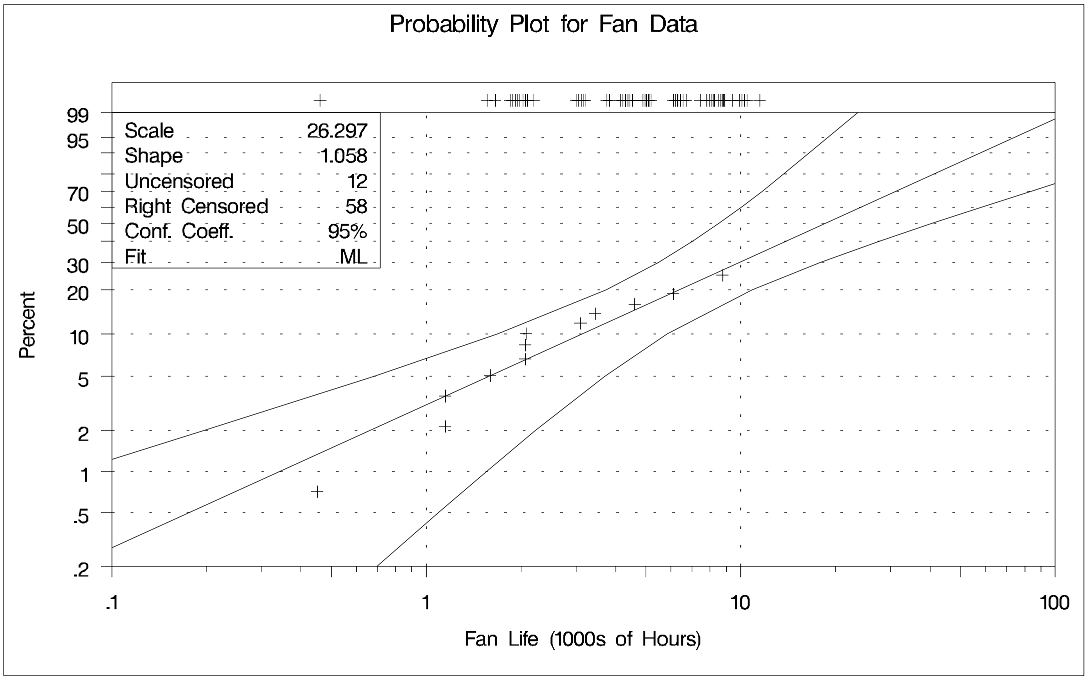

Figure 30.1. Weibull Probability Plot for the Engine Fan Data

The tabular output produced by the preceding SAS statements is shown in Figure 30.2. This consists of summary data, fit information, parameter estimates, distribution percentile estimates, standard errors, and confidence intervals for all estimated quantities.

Model Information

Input Data Set	WORK.FAN	
Analysis Variable	lifetime	Fan Life (1000s of Hours)
Censor Variable	censor	
Distribution	Weibull	
Estimation Method	Maximum Likelihood	
Confidence Coefficient	95%	
Observations Used	70	

Summary of Fit

Observations Used	70
Uncensored Values	12
Right Censored Values	58
Maximum Loglikelihood	-42.248

Weibull Parameter Estimates

			Asymptotic Normal 95% Confidence Limits	
Parameter	Estimate	Standard Error	Lower	Upper
EV Location	3.2694	0.4659	2.3563	4.1826
EV Scale	0.9448	0.2394	0.5749	1.5526
Weibull Scale	26.2968	12.2514	10.5521	65.5344
Weibull Shape	1.0584	0.2683	0.6441	1.7394

Other Weibull Distribution Parameters

Parameter	Value
Mean	25.7156
Mode	1.7039
Median	18.6002

Estimated Covariance Matrix Weibull Parameters

	EV Location	EV Scale
EV Location	0.21705	0.09044
EV Scale	0.09044	0.05733

Estimated Covariance Matrix Weibull Parameters

	Weibull Scale	Weibull Shape
Weibull Scale	150.09724	-2.66446
Weibull Shape	-2.66446	0.07196

Weibull Percentile Estimates

			Asymptotic Normal 95% Confidence Limits	
Percent	Estimate	Standard Error	Lower	Upper
0.1	0.03852697	0.05027782	0.002985	0.49726229
0.2	0.07419554	0.08481353	0.00789519	0.69725757
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
99.9	163.265082	144.264145	28.8905203	922.637827

Figure 30.2. Tabular Output for the Fan Data Analysis

Weibull Analysis Comparing Groups of Data

This example illustrates probability plotting and distribution fitting for data grouped by the levels of a special *group-variable*. The data are from an accelerated life test of an insulating fluid and are the times to electrical breakdown of the fluid under different high voltage levels. Each voltage level defines a subset of data for which a separate analysis and Weibull plot are produced. These data are the 26kV, 30kV, 34kV, and 38kV groups of the data provided by Nelson (1990, p. 129). The following statements create a SAS data set containing the lifetimes and voltages.

```
data fluid;
   input time voltage$ @@;
   datalines;
      5.79 26kV 1579.52 26kV
   2323.70 26kV    7.74 30kV
     17.05 30kV   20.46 30kV
     21.02 30kV   22.66 30kV
     43.40 30kV   47.30 30kV
    139.07 30kV  144.12 30kV
    175.88 30kV  194.90 30kV
      0.19 34kV    0.78 34kV
      0.96 34kV    1.31 34kV
      2.78 34kV    3.16 34kV
      4.15 34kV    4.67 34kV
      4.85 34kV    6.50 34kV
      7.35 34kV    8.01 34kV
      8.27 34kV   12.06 34kV
     31.75 34kV   32.52 34kV
     33.91 34kV   36.71 34kV
     72.89 34kV    0.09 38kV
      0.39 38kV    0.47 38kV
      0.73 38kV    0.74 38kV
      1.13 38kV    1.40 38kV
      2.38 38kV
      ;
run;
```

The variable TIME provides the time to breakdown in minutes, and the variable VOLTAGE provides the voltage level at which the test was conducted. These data are not censored.

The RELIABILITY procedure plots the data for the different voltage levels on the same Weibull probability plot, fits a separate distribution to the data at each voltage level, and superimposes distribution lines on the plot.

The following statements produce the probability plot shown in Figure 30.3 for the variable TIME at each level of the *group-variable* VOLTAGE.

```
proc reliability data=fluid;
   distribution weibull;
   probplot time = voltage / overlay noconf;
run;
```

The input data set FLUID is specified by the DATA= option in the PROC RELIABILITY statement. The PROBPLOT statement option OVERLAY specifies that plots for the groups are to be overlaid rather than displayed separately. The option NOCONF specifies that no confidence bands are to be plotted, since these can interfere with one another on overlaid plots; confidence bands are displayed by default.

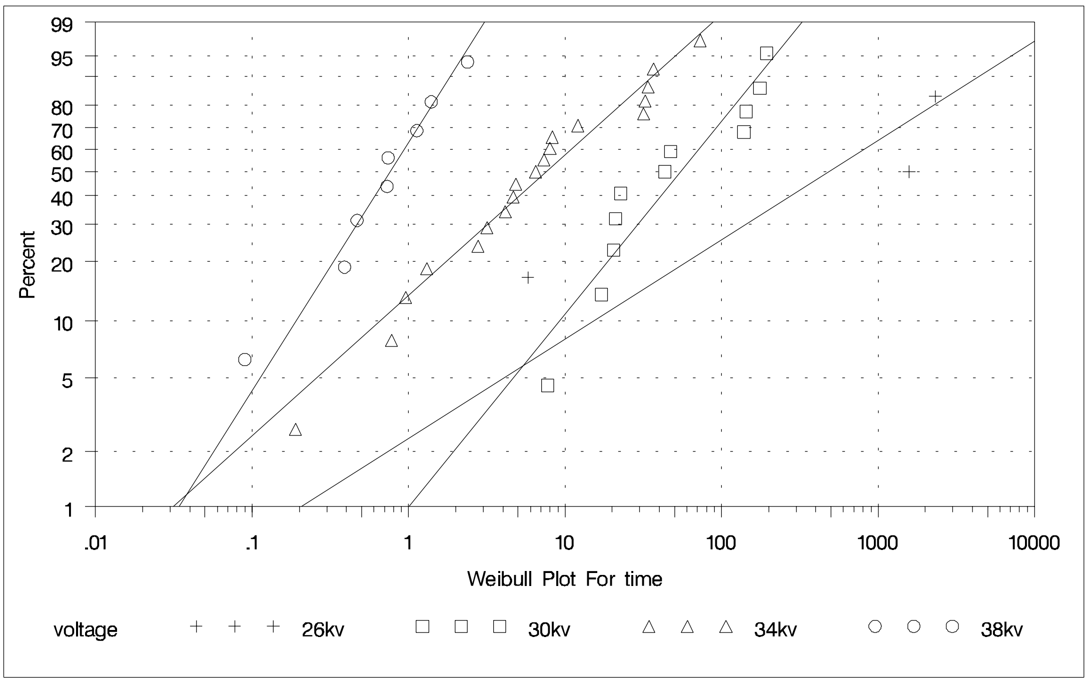

Figure 30.3. Weibull Probability Plot for the Insulating Fluid Data

A summary table that contains information for all groups is displayed. In addition, information identical to that shown in Figure 30.2 is tabulated for each level of voltage. The summary table for all groups and the tables for the 26kV group are shown in Figure 30.4.

```
                           The RELIABILITY Procedure

                         Model Information - All Groups

                  Input Data Set                     WORK.FLUID
                  Analysis Variable                        time
                  Distribution                          Weibull
                  Estimation Method          Maximum Likelihood
                  Confidence Coefficient                    95%
                  Observations Used                          41

                                Summary of Fit
                                                Group

              Observations Used                  3    26kv
              Uncensored Values                  3    26kv
              Maximum Loglikelihood      -6.845551    26kv


                          Weibull Parameter Estimates
                                                   Asymptotic Normal
                                   Standard      95% Confidence Limits
Parameter          Estimate           Error          Lower         Upper   Group

EV Location          6.8625          1.1040         4.6986        9.0264   26kv
EV Scale             1.8342          0.9611         0.6568        5.1226   26kv
Weibull Scale      955.7467       1055.1862       109.7941     8319.6794   26kv
Weibull Shape        0.5452          0.2857         0.1952        1.5226   26kv


                     Other Weibull Distribution Parameters
                     Parameter          Value    Group

                     Mean           1649.4882    26kv
                     Mode              0.0000    26kv
                     Median          487.9547    26kv


                          Weibull Percentile Estimates
                                                  Asymptotic Normal
                                Standard         95% Confidence Limits
 Percent      Estimate             Error          Lower          Upper    Group

     0.1    0.00300636        0.02113841    3.11203E-9     2904.27046    26kv
     0.2    0.01072998        0.06838144    4.03597E-8     2852.65767    26kv
      .          .                 .             .              .          .
      .          .                 .             .              .          .
      .          .                 .             .              .          .
    99.9     33104.172        62018.1074    841.826189     1301796.28    26kv
```

Figure 30.4. Partial Listing of the Tabular Output for the Insulating Fluid Data

Analysis of Accelerated Life Test Data

The following example illustrates the analysis of an accelerated life test for Class-B electrical motor insulation using data provided by Nelson (1990, p. 243). Forty insulation specimens were tested at four temperatures: 150°, 170°, 190°, and 220°C. The purpose of the test is to estimate the median life of the insulation at the design operating temperature of 130°C.

The data are listed in Figure 30.5. Ten specimens of the insulation were tested at each test temperature. The variable TIME provides a specimen time to failure or a censoring time, in hours. The variable CENSOR is equal to 1 if the value of the variable TIME is a right-censoring time and is equal to 0 if the value is a failure time. Some censor times and failure times are identical at some of the temperatures. Rather than repeating identical observations in the input data set, the variable COUNT provides the number of specimens with identical times and temperatures. The variable TEMP provides the test temperature in degrees centigrade. The variable CNTRL is a control variable specifying that percentiles are to be computed only for the first value of TEMP (130°C). The value of TEMP in the first observation (130°C) does not correspond to a test temperature. The missing values in the first observation cause the observation to be excluded from the model fit, and the value of 1 for the variable CNTRL causes percentiles corresponding to a temperature of 130°C to be computed.

Obs	hours	temp	count	censor	intemp	cntrl
1	.	130	.	.	2.48040	1
2	8064	150	10	1	2.36317	0
3	1764	170	1	0	2.25652	0
4	2772	170	1	0	2.25652	0
5	3444	170	1	0	2.25652	0
6	3542	170	1	0	2.25652	0
7	3780	170	1	0	2.25652	0
8	4860	170	1	0	2.25652	0
9	5196	170	1	0	2.25652	0
10	5448	170	3	1	2.25652	0
11	408	190	2	0	2.15908	0
12	1344	190	2	0	2.15908	0
13	1440	190	1	0	2.15908	0
14	1680	190	5	1	2.15908	0
15	408	220	2	0	2.02774	0
16	504	220	3	0	2.02774	0
17	528	220	5	1	2.02774	0

Figure 30.5. Listing of the Class B Insulation Data

An Arrhenius-lognormal model is fitted to the data in this example. In other words, the fitted model has the lognormal (base 10) distribution, and its location parameter μ depends on the centigrade temperature TEMP through the Arrhenius relationship

$$\mu(x) = \beta_0 + \beta_1 x$$

where

$$x = \frac{1000}{TEMP + 273.15}$$

is 1000 times the reciprocal absolute temperature. The lognormal (base e) distribution is also available.

The following SAS statements fit the Arrhenius-lognormal model, and they display the fitted model distributions side-by-side on the probability and the relation plots shown in Figure 30.6.

```
proc reliability data=classb;
   distribution lognormal10;
   freq count;
   model hours*censor(1) = temp /
      relation=arr
      obstats( q=.1 .5 .9 control=cntrl );
   rplot hours*censor(1) = temp /
      pplot
      fit=model
      noconf
      relation = arr
      plotdata
      plotfit 10 50 90
      lupper = 1.e5
      slower=120;
run;
```

The PROC RELIABILITY statement invokes the procedure and specifies CLASSB as the input data set. The DISTRIBUTION statement specifies that the lognormal (base 10) distribution is to be used for maximum likelihood parameter estimation and probability plotting. The FREQ statement specifies that the variable COUNT is to be used as a frequency variable; that is, if COUNT=n, then there are n specimens with the time and temperature specified in the observation.

The MODEL statement fits a linear regression equation for the distribution location parameter as a function of independent variables. In this case, the MODEL statement also transforms the independent variable through the Arrhenius relationship. The dependent variable is specified as TIME. A value of 1 for the variable CENSOR indicates that the corresponding value of TIME is a right-censored observation; otherwise, the value is a failure time. The temperature variable TEMP is specified as the independent variable in the model. The MODEL statement option RELATION=ARR specifies the Arrhenius relationship.

The option OBSTATS requests observation-wise statistics. The options in parentheses following OBSTATS indicate which statistics are to be computed. In this case, QUANTILE = .1 .5 .9 specifies that quantiles of the fitted distribution are to be computed for the value of the variable TEMP at each observation. The CONTROL= option requests quantiles only for those observations in which the variable CNTRL has a value of 1. This eliminates unnecessary quantiles in the OBSTATS table since, in this case, only the quantiles at the design temperature of 130°C are of interest.

The RPLOT, or RELATIONPLOT, statement displays a plot of the lifetime data and the fitted model. The dependent variable TIME, the independent variable TEMP, and the censoring indicator CENSOR are the same as in the MODEL statement. The option FIT=MODEL specifies that the model fitted with the preceding MODEL statement is to be used for probability plotting and in the relation plot. The option RELATION=ARR specifies an Arrhenius scale for the horizontal axis of the relation plot. The PPLOT option specifies that a probability plot is to be displayed alongside the relation plot. The type of probability plot is determined by the distribution named in the DISTRIBUTION statement, in this case, a lognormal (base 10) distribution. Weibull, extreme value, lognormal (base e), normal, loglogistic, and logistic distributions are also available. The NOCONF option suppresses the default percentile confidence bands on the probability plot. The PLOTDATA option specifies that the failure times are to be plotted on the relation plot. The PLOTFIT option specifies that the 10th, 50th, and 90th percentiles of the fitted relationship are to be plotted on the relation plot. The options LUPPER and SLOWER specify an upper limit on the life axis scale and a lower limit on the stress (temperature) axis scale in the plots.

The plots produced by the preceding statements are shown in Figure 30.6. The plot on the left is an overlaid lognormal probability plot of the data and the fitted model. The plot on the right is a relation plot showing the data and the fitted relation. The fitted straight lines are percentiles of the fitted distribution at each temperature. An Arrhenius relation fitted to the data, plotted on an Arrhenius plot, yields straight percentile lines.

Since all the data at 150°C are right censored, there are no failures corresponding to 150°C on the probability plot. However, the fitted distribution at 150°C is plotted on the probability plot.

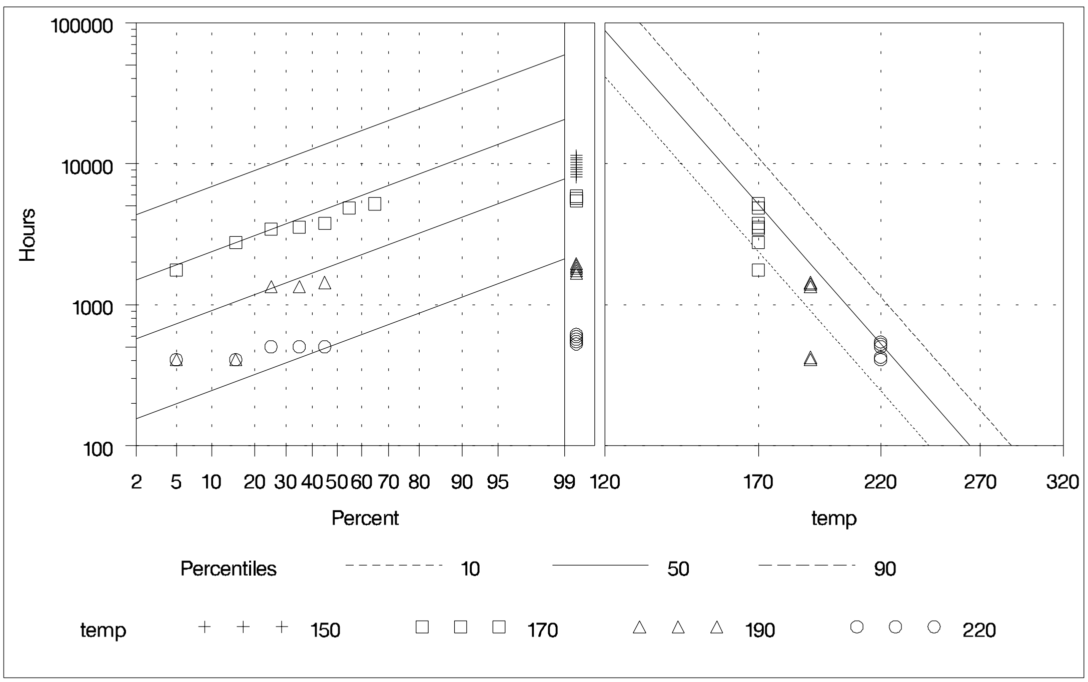

Figure 30.6. Probability and Relation Plots for the Class B Insulation Data

The tabular output requested with the MODEL statement is shown in Figure 30.7. The “Model Information” table provides general information about the data and model. The “Summary of Fit” table shows the number of observations used, the number of failures and of censored values (accounting for the frequency count), and the maximum log likelihood for the fitted model.

The “Lognormal Parameter Estimates” table contains the Arrhenius-lognormal model parameter estimates, their standard errors, and confidence interval estimates. In this table, INTERCEPT is the maximum likelihood estimate of β_0, TEMP is the estimate of β_1, and Scale is the estimate of the lognormal scale parameter, σ.

```
                       The RELIABILITY Procedure

                           Model Information

     Input Data Set                 WORK.CLASSB
     Analysis Variable                    hours    Hours
     Relation                         Arrhenius
     Censor Variable                     censor
     Frequency Variable                   count
     Distribution              Lognormal (Base 10)

                             Summary of Fit

                    Observations Used                  16
                    Uncensored Values                  17
                    Right Censored Values              23
                    Missing Observations                1
                    Maximum Loglikelihood       -12.96533


                     Lognormal Parameter Estimates
                                                Asymptotic Normal
                                    Standard      95% Confidence Limits
       Parameter      Estimate         Error         Lower         Upper

       Intercept       -6.0182        0.9467       -7.8737       -4.1628
       temp             4.3103        0.4366        3.4546        5.1660
       Scale            0.2592        0.0473        0.1812        0.3708


                         Observation Statistics
             Hours       censor          temp         count         PCNTL

                 .            .           130             .     21937.658
                 .            .           130             .     47135.132
                 .            .           130             .     101274.29

                         Observation Statistics
                    Hours       STDERR         LOWER         UPPER

                        .     6959.151     11780.636     40851.857
                        .    16125.548     24106.685     92162.016
                        .      42061.1     44872.401     228569.92
```

Figure 30.7. MODEL Statement Output for the Class B Data

The “Observation Statistics” table provides the estimates of the fitted distribution quantiles, their standard errors, and confidence limits. These are given only for the value of 130°C, as specified with the CONTROL= option in the MODEL statement.

The predicted median life at 130°C corresponds to a quantile of 0.5, and it is approximately 47,134 hours.

In addition to the MODEL statement output in Figure 30.7, the RELIABILITY procedure produces tabular output for each temperature that is identical to the output produced with the PROBPLOT statement. This output is not shown here.

Weibull Analysis of Interval Data with Common Inspection Schedule

Table 30.1 shows data for 167 identical turbine parts provided by Nelson (1982, p. 415). The parts were inspected at certain times to determine which parts had cracked since the last inspection. The times at which parts develop cracks are to be fitted with a Weibull distribution.

Table 30.1. Turbine Part Cracking Data

Inspection (Months)		Number	
Start	End	Cracked	Cumulative
0	6.12	5	5
6.12	19.92	16	21
19.92	29.64	12	33
29.64	35.40	18	51
35.40	39.72	18	69
39.72	45.24	2	71
45.24	52.32	6	77
52.32	63.48	17	94
63.48	Survived	73	167

Table 30.1 shows the time in months of each inspection period and the number of cracked parts found in each period. These data are said to be interval censored since only the time interval in which failures occurred is known, not the exact failure times. Seventy-three parts had not cracked at the last inspection, which took place at 63.48 months. These 73 lifetimes are right censored, since the lifetimes are known only to be greater than 63.48 months.

The interval data in this example is read from a SAS data set with a special structure. All units must have a common inspection schedule. This type of interval data is called readout data. The SAS data set named CRACKS, shown in Figure 30.8, provides the data in Table 30.1 with this structure. The variable TIME is the inspection time, that is, the upper endpoint of each interval. The variable UNITS is the number of unfailed units at the beginning of each interval, and the variable FAIL is the number of units with cracks at the inspection time.

Obs	time	units	fail
1	6.12	167	5
2	19.92	162	16
3	29.64	146	12
4	35.40	134	18
5	39.72	116	18
6	45.24	98	2
7	52.32	96	6
8	63.48	90	17

Figure 30.8. Listing of the Turbine Part Cracking Data

The following statements use the RELIABILITY procedure to produce the probability plot in Figure 30.9 for the data in the data set CRACKS.

```
proc reliability data=cracks;
   freq fail;
   nenter units;
   distribution weibull;
   probplot time / readout
                   pconfplt
                   noconf;
run;
```

The FREQ statement specifies that the variable FAIL provides the number of failures in each interval. The NENTER statement specifies that the variable UNITS provides the number of unfailed units at the beginning of each interval. The DISTRIBUTION statement specifies that the Weibull distribution is used for parameter estimation and probability plotting. The PROBPLOT statement requests a probability plot of the data.

The PROBPLOT statement option READOUT indicates that the data in the CRACKS data set are readout (or interval) data. The option PCONFPLT specifies that confidence intervals for the cumulative probability of failure are to be plotted. The confidence intervals for the cumulative probability are based on the binomial distribution for time intervals until right censoring occurs. For time intervals after right censoring occurs, the binomial distribution is not valid, and a normal approximation is used to compute confidence intervals.

The option NOCONF suppresses the display of confidence intervals for distribution percentiles in the probability plot.

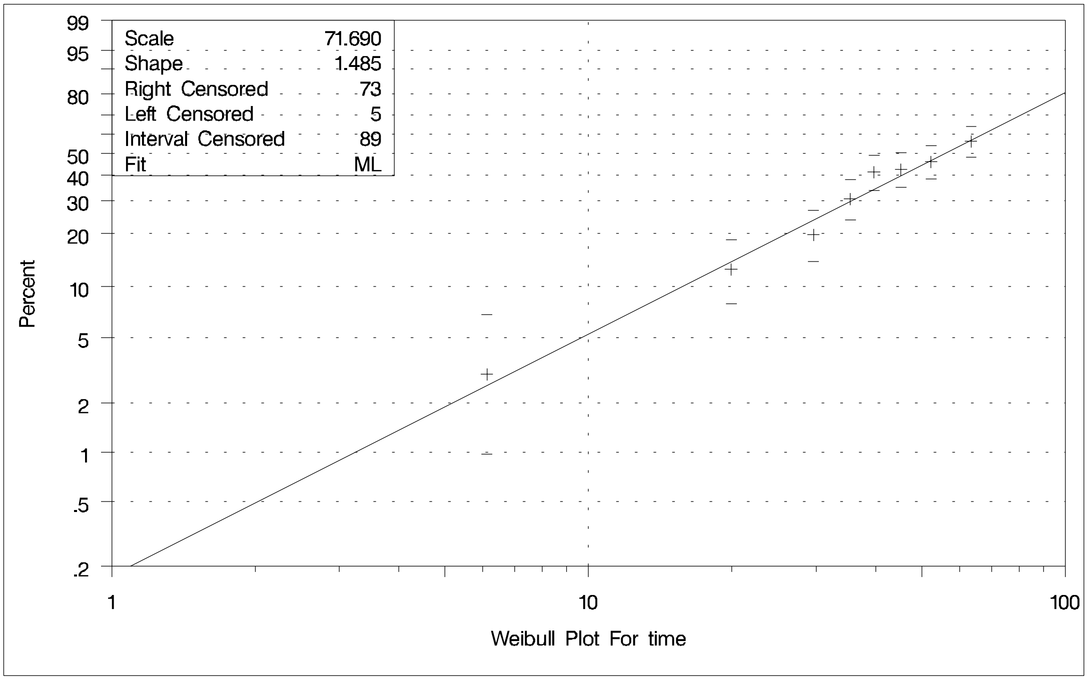

Figure 30.9. Weibull Probability Plot for the Part Cracking Data

A partial listing of the tabular output produced by the preceding SAS statements is shown in Figure 30.10. By default, the specified Weibull distribution is fitted by maximum likelihood. The line plotted on the probability plot and the tabular output summarize this fit. For interval data, the estimated cumulative probabilities and associated confidence intervals are tabulated. In addition, general fit information, parameter estimates, percentile estimates, standard errors, and confidence intervals are tabulated.

```
                         Model Information

             Input Data Set                   WORK.CRACKS
             Analysis Variable                       time
             Frequency Variable                      fail
             NENTER Variable                        units
             Distribution                         Weibull
             Estimation Method         Maximum Likelihood
             Confidence Coefficient                   95%
             Observations Used                          8


                   Cumulative Probability Estimates
                                         95% Confidence
   Lower       Upper     Cumulative          Limits          Standard
 Lifetime    Lifetime    Probability     Lower      Upper       Error

        .        6.12         0.0299    0.0098     0.0685      0.0132
     6.12       19.92         0.1257    0.0796     0.1858      0.0257
    19.92       29.64         0.1976    0.1401     0.2662      0.0308
    29.64        35.4         0.3054    0.2366     0.3813      0.0356
     35.4       39.72         0.4132    0.3376     0.4918      0.0381
    39.72       45.24         0.4251    0.3491     0.5039      0.0383
    45.24       52.32         0.4611    0.3838     0.5398      0.0386
    52.32       63.48         0.5629    0.4841     0.6394      0.0384

                         Summary of Fit

                Observations Used                  8
                Right Censored Values             73
                Left Censored Values               5
                Interval Censored Values          89
                Maximum Loglikelihood      -309.6684


                    Weibull Parameter Estimates
                                                Asymptotic Normal
                                    Standard   95% Confidence Limits
Parameter           Estimate           Error        Lower        Upper

EV Location           4.2724          0.0744       4.1265       4.4182
EV Scale              0.6732          0.0664       0.5549       0.8168
Weibull Scale        71.6904          5.3335      61.9634      82.9444
Weibull Shape         1.4854          0.1465       1.2242       1.8022


                           Other Weibull
                      Distribution Parameters
                      Parameter          Value

                      Mean             64.7966
                      Mode             33.7622
                      Median           56.0144


                    Weibull Percentile Estimates
                                              Asymptotic Normal
                                  Standard   95% Confidence Limits
   Percent      Estimate             Error        Lower        Upper

       0.1    0.68534385    0.29999861    0.29060848    1.61625083
       0.2    1.09324674    0.42889777    0.50673224     2.3586193
        .         .              .             .             .
        .         .              .             .             .
        .         .              .             .             .
      99.9    263.348102    44.7205513    188.791789    367.347666
```

Figure 30.10. Partial Listing of the Tabular Output for the Part Cracking Data

In this example, the number of unfailed units at the beginning of an interval minus the number failing in the interval is equal to the number of unfailed units entering the next interval. This is not always the case since some unfailed units might be removed from the test at the end of an interval; that is, they might be right censored. The special structure of the input SAS data set required for interval data enables the RELIABILITY procedure to analyze this more general case.

Lognormal Analysis with Arbitrary Censoring

This example illustrates analyzing data that have more general censoring than in the previous example. The data can be a combination of exact failure times, left censored, right censored, and interval censored data. The intervals can be overlapping, unlike in the previous exmaple, where the interval endpoints had to be the same for all units.

Table 30.2 shows data from Nelson (1982, p. 409), analyzed by Meeker and Escobar (1998, p. 135). Each of 435 turbine wheels was inspected once to determine whether a crack had developed in the wheel or not. The inspection time (in 100s of hours), the number inspected at the time that had cracked, and the number not cracked are shown in the table. The quantity of interest is the time for a crack to develop.

Table 30.2. Turbine Wheel Cracking Data

Inspection Time (100 hours)	Number Cracked	Number Not Cracked
4	0	39
10	4	49
14	2	31
18	7	66
22	5	25
26	9	30
30	9	33
34	6	7
38	22	12
42	21	19
46	21	15

These data consist only of left and right censored lifetimes. If a unit has developed a crack at an inspection time, the unit is left-censored at the time; if a unit has not developed a crack, it is right-censored at the time. For example, there are 4 left-censored lifetimes and 49 right-censored lifetimes at 1000 hours.

The following statements create a SAS data set named TURBINE thet contains the data in the format necessary for analysis by the RELIABILITY procedure.

```
data turbine;
   label t1 = 'Time of Cracking (Hours x 100 )';
   input t1 t2 f;
   datalines;
.   4  0
4   .  39
.  10  4
10  .  49
.  14  2
14  .  31
.  18  7
18  .  66
.  22  5
22  .  25
.  26  9
26  .  30
.  30  9
30  .  33
.  34  6
34  .  7
.  38  22
38  .  12
.  42  21
42  .  19
.  46  21
46  .  15
;
run;
```

The variables T1 and T2 represent the inspection times and determine whether the observation is right or left censored. If T1 is missing (.), then T2 represents a left-censoring time; if T2 is missing, T1 represents a right-censoring time. The variable F is the number of units that were found to be cracked for left-censored observations, or not cracked for right-censored observations at an inspection time.

The following statements use the RELIABILITY procedure to produce the probability plot in Figure 30.11 for the data in the data set TURBINE.

```
proc reliability data = turbine;
   distribution lognormal;
   freq f;
   pplot ( t1 t2 ) / maxitem = 5000
                     ppout ;

   run;
```

The DISTRIBUTION statement specifies that a lognormal probability plot be created. The FREQ statement identifies the frequency variable F. The option MAXITEM = 5000 specifies that the iterative algorithm that computes the points on the probability plot can take a maximum of 5000 iterations. The algorithm does not converge for this data in the default 1000 iterations, so the maximum number of iterations needs to be increased for convergence. The option PPOUT specifies that a table of the cumulative probabilities plotted on the probability plot be printed, along with standard errors and confidence limits.

The tabular output for the maximum likelihood lognormal fit for this data is shown in Figure 30.12. Figure 30.11 shows the resulting lognormal probability plot with the computed cumulative probability estimates and the lognormal fit line.

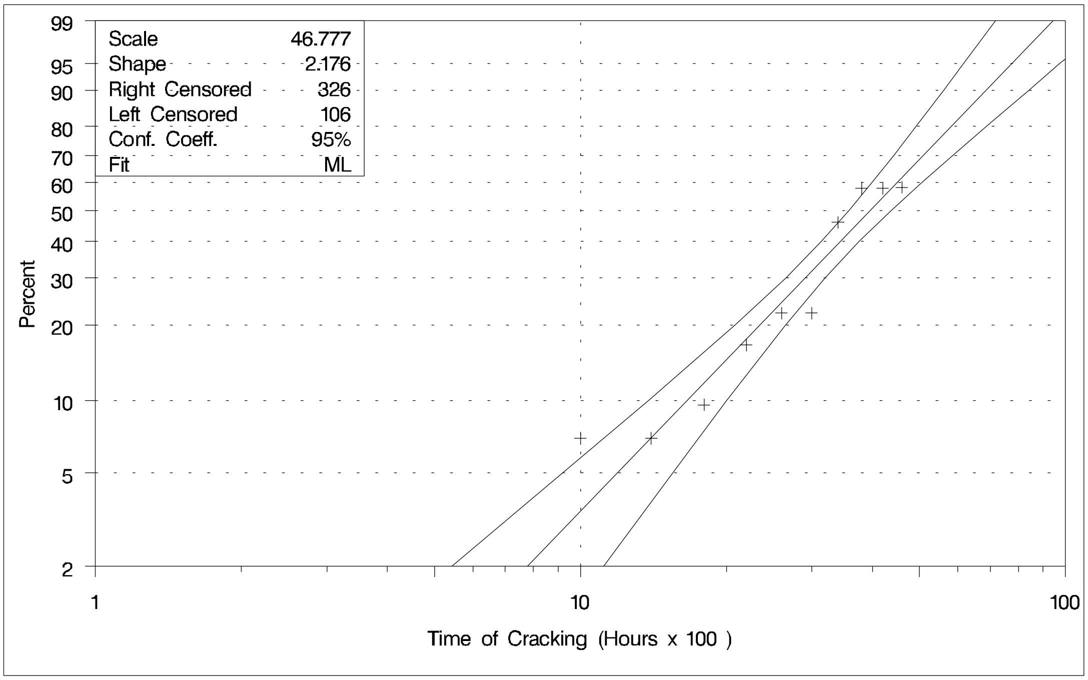

Figure 30.11. Lognormal Probability Plot for the Turbine Wheel Data

```
                          Model Information

   Input Data Set                  WORK.TURBINE
   Analysis Variable                         t1    Time of Cracking
                                                   (Hours x 100 )
   Analysis Variable                         t2
   Frequency Variable                         f
   Distribution                 Lognormal (Base e)
   Estimation Method            Maximum Likelihood
   Confidence Coefficient                   95%
   Observations Used                         21

                    Cumulative Probability Estimates
                                          95% Confidence
      Lower       Upper     Cumulative        Limits         Standard
   Lifetime    Lifetime    Probability     Lower      Upper       Error

          .           4         0.0000    0.0000     0.0000      0.0000
         10          10         0.0698    0.0264     0.1720      0.0337
         14          14         0.0698    0.0177     0.2384      0.0473
         18          18         0.0959    0.0464     0.1878      0.0345
         22          22         0.1667    0.0711     0.3432      0.0680
         26          26         0.2222    0.1195     0.3757      0.0657
         30          30         0.2222    0.1203     0.3738      0.0650
         34          34         0.4615    0.2236     0.7184      0.1383
         38          38         0.5809    0.4085     0.7356      0.0865
         42          42         0.5809    0.4280     0.7198      0.0766
         46          46         0.5836    0.4195     0.7311      0.0822

                           Summary of Fit

                  Observations Used                 21
                  Uncensored Values                  0
                  Right Censored Values            326
                  Left Censored Values             106
                  Maximum Loglikelihood      -190.7315

                      Lognormal Parameter Estimates
                                                  Asymptotic Normal
                                   Standard      95% Confidence Limits
       Parameter       Estimate        Error          Lower         Upper

       Location          3.6999       0.0708         3.5611        3.8387
       Scale             0.7199       0.0887         0.5655        0.9165

                             Other Lognormal
                         Distribution Parameters
                         Parameter          Value

                         Mean             52.4062
                         Mode             24.0870
                         Median           40.4436

                      Lognormal Percentile Estimates
                                                 Asymptotic Normal
                                  Standard      95% Confidence Limits
         Percent      Estimate        Error          Lower         Upper

             0.1    4.37231983   1.01951851     2.76842301     6.9054406
             0.2    5.09347486   1.09461144     3.34261991    7.76142271
              .          .            .              .             .
              .          .            .              .             .
              .          .            .              .             .
            99.9    374.099407   121.716176     197.716048    707.835138
```

Figure 30.12. Partial Listing of the Tabular Output for the Turbine Wheel Data

Regression Modeling

This example is an illustration of a Weibull regression model using a load accelerated life test of rolling bearings, with data provided by Nelson (1990, p. 305). Bearings are tested at four different loads, and lifetimes in 10^6 of revolutions are measured. The data are shown in Table 30.3. An outlier identified by Nelson (1990) is omitted.

Table 30.3. Bearing Lifetime Data

Load	Life (10^6 Revolutions)									
0.87	1.67	2.2	2.51	3.00	3.90	4.70	7.53	14.7	27.76	37.4
0.99	0.80	1.0	1.37	2.25	2.95	3.70	6.07	6.65	7.05	7.37
1.09	0.18	0.2	0.24	0.26	0.32	0.32	0.42	0.44	0.88	
1.18	0.073	0.098	0.117	0.135	0.175	0.262	0.270	0.350	0.386	0.456

These data are modeled with a Weibull regression model in which the independent variable is the logarithm of the load. The model is

$$\mu_i = \beta_0 + \beta_1 x_i$$

where μ_i is the location parameter of the extreme value distribution and

$$x_i = \log(\text{load})$$

for the ith bearing. The following statements create and list a SAS data set containing the loads, log loads, and bearing lifetimes.

```
data bearing;
   input load life;
   lload = log(load);
   datalines;
   0.87 1.67
   0.87 2.2
   .
   .
   .
   1.18 .456
   ;

proc print data=bearing;
run;
```

Figure 30.13 shows a partial listing of the bearing data.

```
                     Obs    load     life       lload

                       1    0.87     1.670    -0.13926
                       2    0.87     2.200    -0.13926
                       3    0.87     2.510    -0.13926
                       .      .        .         .
                       .      .        .         .
                       .      .        .         .
                      39    1.18     0.456     0.16551
```

Figure 30.13. Partial Listing of the Bearing Data

The following statements fit the regression model by maximum likelihood using the Weibull distribution.

```
ods output modobstats = RESIDUAL;
proc reliability data=bearing;
   distribution weibull;
   model life = lload / covb
                        corrb
                        obstats
                        ;
run;

proc print data=RESIDUAL;
run;
```

The PROC RELIABILITY statement invokes the procedure and identifies BEARING as the input data set. The DISTRIBUTION statement specifies the Weibull distribution for model fitting. The MODEL statement specifies the regression model, identifying LIFE as the variable that provides the response values (the lifetimes) and LLOAD as the independent variable (the log loads). The MODEL statement option COVB requests the regression parameter covariance matrix, and the CORRB option requests the correlation matrix. The option OBSTATS requests a table that contains residuals, predicted values, and other statistics. The ODS output statement creates a SAS data set named RESIDUAL that contains the table created by the OBSTATS option.

Figure 30.14 shows the tabular output produced by the RELIABILITY procedure. The "Weibull Parameter Estimates" table contains parameter estimates, their standard errors, and 95% confidence intervals. In this table, INTERCEPT corresponds to β_0, LLOAD corresponds to β_1, and SHAPE corresponds to the Weibull shape parameter. Figure 30.15 shows a partial listing of the output data set RESIDUAL.

```
                          The RELIABILITY Procedure

                              Model Information

                      Input Data Set       WORK.BEARING
                      Analysis Variable            life
                      Distribution              Weibull

                            Parameter Information

                          PRM1           Intercept
                          PRM2           lload
                          PRM3           EV Scale

                               Summary of Fit

                    Observations Used                  39
                    Uncensored Values                  39
                    Maximum Loglikelihood       -51.77737


                         Weibull Parameter Estimates
                                                  Asymptotic Normal
                                       Standard  95% Confidence Limits
     Parameter          Estimate        Error        Lower         Upper

     Intercept            0.8323       0.1410       0.5560        1.1086
     lload              -13.8529       1.2333     -16.2703      -11.4356
     EV Scale             0.8043       0.0999       0.6304        1.0260
     Weibull Shape        1.2434       0.1545       0.9746        1.5862


                         Estimated Covariance Matrix
                              Weibull Parameters

                           PRM1              PRM2              PRM3

           PRM1         0.01987          -0.04374          -0.00492
           PRM2        -0.04374           1.52113           0.01578
           PRM3        -0.00492           0.01578           0.00999


                         Estimated Correlation Matrix
                              Weibull Parameters

                           PRM1              PRM2              PRM3

           PRM1          1.0000           -0.2516           -0.3491
           PRM2         -0.2516            1.0000            0.1281
           PRM3         -0.3491            0.1281            1.0000
```

Figure 30.14. Analysis Results for the Bearing Data

Obs	life	lload	XBETA	SURV	RESID	SRESID	ARESID
1	1.67	-0.139262	2.7614742	0.9407681	-2.248651	-2.795921	-2.795921
2	2.2	-0.139262	2.7614742	0.9175782	-1.973017	-2.453205	-2.453205
3	2.51	-0.139262	2.7614742	0.9036277	-1.841191	-2.289296	-2.289296
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
39	0.456	0.1655144	-1.460578	0.0987061	0.6753158	0.8396724	0.8396724

Figure 30.15. Partial Listing of RESIDUAL

The value of the lifetime LIFE and the log load LLOAD are included in this data set, as well as statistics computed from the fitted model. The variable _XBETA_ is the value of the linear predictor

$$\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \mathrm{LLOAD}\hat{\beta}_1$$

for each observation. The variable _SURV_ contains the value of the reliability function, the variable _SRESID_ contains the standardized residual, and the variable _ARESID_ contains a residual adjusted for right-censored observations. Since there are no censored values in these data, _SRESID_ is equal to _ARESID_ for all the bearings. See Table 30.21 and Table 30.22 for other statistics that are available in the OBSTATS table and data set. See the section "Regression Model Observation-Wise Statistics" on page 1017 for a description of the residuals and other statistics.

If the fitted regression model is adequate, the standardized residuals have a standard extreme value distribution. You can check the residuals by creating an extreme value probability plot of the residuals using the RELIABILITY procedure and the RESIDUAL data set. The following statements create the plot in Figure 30.16.

```
proc reliability data=residual;
   distribution ev;
   probplot sresid;
run;
```

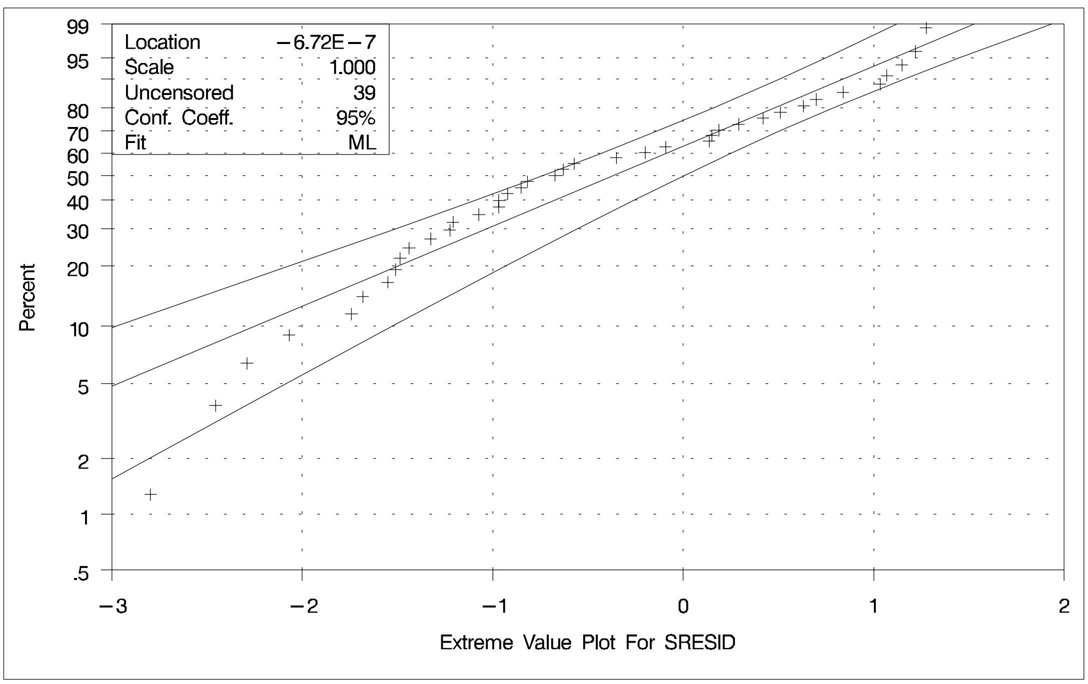

Figure 30.16. Extreme Value Probability Plot for the Standardized Residuals

Although the estimated location is near zero and the estimated scale is near one, the plot reveals systematic curvature, indicating that the Weibull regression model might be inadequate.

Analysis of Recurrence Data on Repairs

This example illustrates analysis of recurrence data from repairable systems. Repair data analysis differs from life data analysis, where units fail only once. As a repairable system ages, it accumulates repairs and costs of repairs. The RELIABILITY procedure provides a nonparametric estimate and plot of the *mean cumulative function* (MCF) for the number or cost of repairs for a population of repairable systems.

The nonparametric estimate of the MCF, the variance of the MCF estimate, and confidence limits for the MCF estimate are based on the work of Nelson (1995). The MCF, also written as $M(t)$, is defined by Nelson (1995) to be the *population mean* of the distribution of the cumulative number or cost of repairs at age t. The method does not assume any underlying structure for the repair process.

Figure 30.17 is a partial listing of the SAS data set VALVE, which contains repair histories of 41 diesel engines in a fleet (Nelson 1995). The valve seats in these engines wear out and must be replaced. The variable ID is a unique identifier for individual engines. The variable DAYS provides the engine age in days. The value of the variable VALUE is 1 if the age is a valve seat replacement age or -1 if the age is the end of history, or censoring age, for the engine.

Obs	id	days	value
1	251	761	-1
2	252	759	-1
3	327	98	1
4	327	667	-1
5	328	326	1
6	328	653	1
7	328	653	1
8	328	667	-1
9	329	665	-1
.	.	.	.
.	.	.	.
.	.	.	.
80	416	202	1
81	416	563	1
82	416	570	1
83	416	585	-1
84	417	587	-1
85	418	578	-1
86	419	578	-1
87	420	586	-1
88	421	585	-1
89	422	582	-1

Figure 30.17. Partial Listing of the Valve Seat Data

The following statements produce the graphical display in Figure 30.18.

```
proc reliability data=valve;
   unitid id;
   mcfplot days*value( -1 );
run;
```

The UNITID statement specifies that the variable ID uniquely identifies each system. The MCFPLOT statement requests a plot of the MCF estimates as a function of the age variable DAYS, and it specifies -1 as the value of the variable VALUE, which identifies the end of history for each engine (system).

In Figure 30.18, the MCF estimates and confidence limits are plotted versus system age in days. The end-of-history ages are plotted in an area at the top of the plot. Except for the last few points, the plot is essentially a straight line, suggesting a constant replacement rate. Consequently, the prediction of future replacements of valve seats can be based on a fitted line in this case.

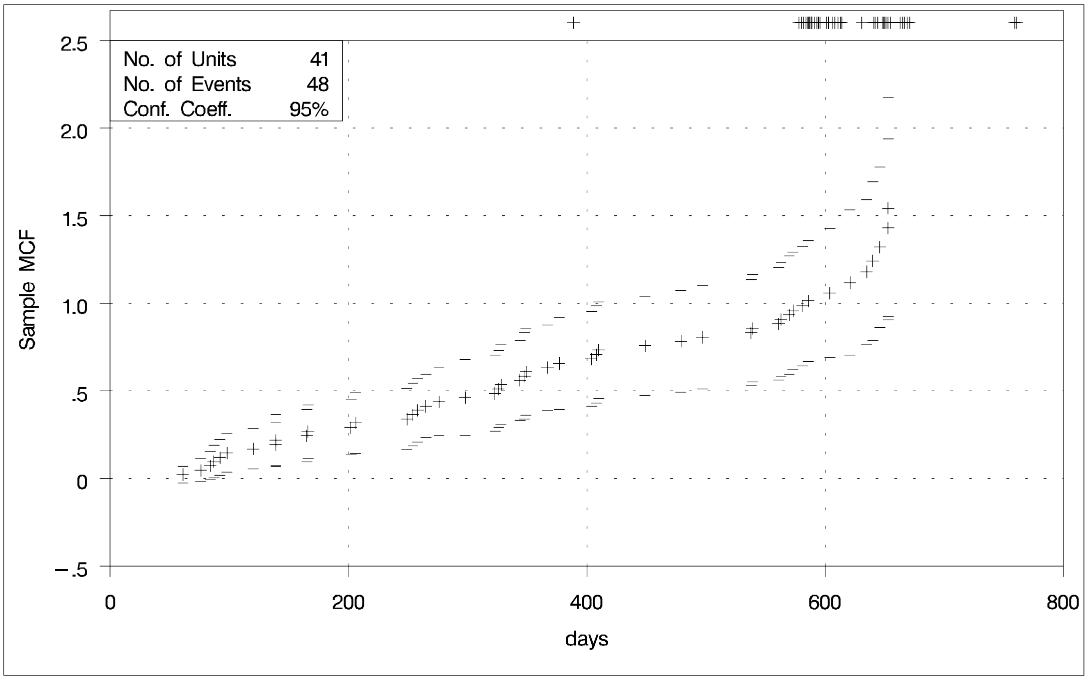

Figure 30.18. Mean Cumulative Function for the Number of Repairs

A partial listing of the tabular output is shown in Figure 30.19. It contains a summary of the repair data, estimates of the MCF, the Nelson (1995) standard errors, and confidence intervals for the MCF. The MCF estimates, standard errors, and confidence limits are shown as missing values (.) at the end of history points, since they are not computed at these points.

```
                        The RELIABILITY Procedure

                           Repair Data Summary

                  Input Data Set              WORK.VALVE
                  Observations Used           89
                  Number of Units             41
                  Number of Events            48


                           Repair Data Analysis

                   Sample    Standard   95% Confidence Limits    Unit
         Age          MCF       Error       Lower       Upper    ID

       61.00        0.024       0.024      -0.023       0.072    393
       76.00        0.049       0.034      -0.018       0.116    395
       84.00        0.073       0.041      -0.008       0.154    330
          .           .           .           .           .        .
          .           .           .           .           .        .
          .           .           .           .           .        .
      667.00          .           .           .           .      327
      759.00          .           .           .           .      252
      761.00          .           .           .           .      251
```

Figure 30.19. Partial Listing of the Output for the Valve Seat Data

Parametric modeling of the repair process requires more assumptions than nonparametric modeling, and considerable work has been done in this area. Ascher and Feingold (1984) describe parametric models for repair processes. For example, repairs are sometimes modeled as a nonhomogeneous Poisson process. The current release of the RELIABILITY procedure does not include this type of parametric modeling, although it is planned for future releases. The MCF plot might be a first step in modeling a repair process, but, in many cases, it provides the required answers without further analysis. An estimate of the MCF for a sample of systems aids engineers in determining the repair rate at any age and the increase or decrease of repair rate with population age. The estimate is also useful for predicting the number of future repairs.

Comparison of Two Samples of Repair Data

Nelson (1995) and Doganaksoy and Nelson (1991) show how the difference of MCFs from two samples can be used to compare the populations from which they are drawn. The RELIABILITY procedure provides Doganaksoy and Nelson's confidence intervals for the pointwise difference of the two MCFs, which can be used to assess whether the difference is statistically significant.

Doganaksoy and Nelson (1991) give an example of two samples of locomotives with braking grids from two different production batches. Figure 30.20 is a partial listing of the data. The variable ID is a unique identifier for individual locomotives. The variable DAYS provides the locomotive age in days. The variable VALUE is 1 if the age corresponds to a valve seat replacement or -1 if the age corresponds to the locomotive's latest age (the current end of its history). The variable SAMPLE is a group variable that identifies the grid production batch.

Obs	sample	ID	days	value
1	Sample1	S1-01	462	1
2	Sample1	S1-01	730	-1
3	Sample1	S1-02	364	1
4	Sample1	S1-02	391	1
5	Sample1	S1-02	548	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
82	Sample2	S2-18	264	1
83	Sample2	S2-18	415	-1

Figure 30.20. Listing of the Braking Grids Data

The following statements request the Nelson (1995) nonparametric estimate and confidence limits for the difference of the MCF functions shown in Figure 30.21 for the braking grids.

```
proc reliability data=grids;
   unitid ID;
   mcfplot days*value(-1) =  sample  / mcfdiff;
run;
```

The MCFPLOT statement requests a plot of each MCF estimate as a function of age (provided by DAYS), and it specifies that the end of history for each system is identified by VALUE equal to -1. The variable SAMPLE identifies the two samples of braking grids. The option MCFDIFF requests that the difference between the MCFs of the two groups given in the variable SAMPLE be computed and plotted. Confidence limits for the MCF difference are also computed and plotted. The UNITID statement specifies that the variable ID uniquely identifies each system.

Figure 30.21 shows the plot of the MCF difference function and pointwise 95% confidence intervals. Since the pointwise confidence limits do not include zero for some system ages, the difference between the two populations is statistically significant.

A partial listing of the tabular output is shown in Figure 30.22. It contains a summary of the repair data for the two samples, estimates, standard errors, and confidence intervals for the MCF difference.

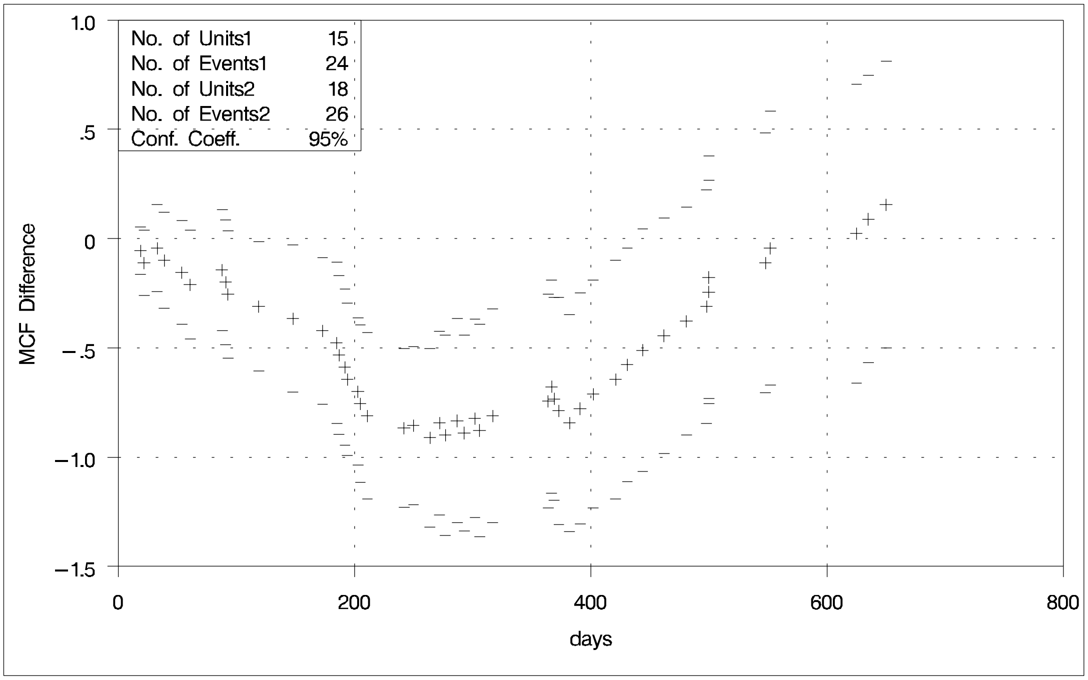

Figure 30.21. Mean Cumulative Function Difference

```
                    The RELIABILITY Procedure

                   MCF Difference Data Summary

              Input Data Set              WORK.GRIDS
              Group 1                     Sample1
              Observations Used           39
              Number of Units             15
              Number of Events            24
              Group 2                     Sample2
              Observations Used           44
              Number of Units             18
              Number of Events            26


                     Sample MCF Differences
                                        95% Confidence
                   MCF     Standard         Limits            Unit
     Age    Difference        Error     Lower      Upper     ID

   19.00        -0.056        0.056    -0.164      0.053     S2-16
   22.00        -0.111        0.076    -0.261      0.038     S2-12
   33.00        -0.044        0.101    -0.243      0.154     S1-13
     .             .            .         .          .          .
     .             .            .         .          .          .
     .             .            .         .          .          .
  635.00         0.089        0.336    -0.569      0.747     S1-11
  650.00         0.156        0.335    -0.500      0.811     S1-11
```

Figure 30.22. Partial Listing of the Output for the Braking Grids Data

Analysis of Binomial Data

This example illustrates the analysis of binomial proportions using capacitor failure data from nine circuit boards given by Nelson (1982, p. 451). The following statements create and list a SAS data set named BINEX containing the data.

```
data binex;
   input board sample fail;
   datalines;
   1 84 2
   2 72 3
   3 72 5
   4 119 19
   5 538 21
   6 51 2
   7 517 9
   8 462 18
   9 143 2
   ;

proc print data=binex;
run;
```

Figure 30.23 displays a listing of the data. The variable BOARD identifies the circuit board, the variable SAMPLE provides the number of capacitors on the boards, and the variable FAIL provides the number of capacitors failing on the boards.

Obs	board	sample	fail
1	1	84	2
2	2	72	3
3	3	72	5
4	4	119	19
5	5	538	21
6	6	51	2
7	7	517	9
8	8	462	18
9	9	143	2

Figure 30.23. Listing of the Capacitor Data

The following statements analyze the proportion of capacitors failing.

```
proc reliability;
   distribution binomial;
   analyze fail(sample) = board / predict(1000)
                                  tolerance(.05);
run;
```

The DISTRIBUTION statement specifies the binomial distribution. The analysis requested with the ANALYZE statement consists of tabular output only. Graphical output is not available for the binomial distribution. The variable FAIL provides the number of capacitors failing on each board, the variable SAMPLE provides the sample size (number of capacitors) for each board, and the variable BOARD identifies the individual boards. The statement option PREDICT(1000) requests the predicted number of capacotors failing and prediction limits in a future sample of size 1000. The option TOLERANCE(.05) requests the sample size required to estimate the binomial proportion to within 0.05. Figure 30.24 displays the results of the analysis.

The "Pooled Data Analysis" table displays the estimated binomial probability and exact binomial confidence limits when data from all boards are pooled. The chi-squared value and *p*-value for a test of equality of the binomial probabilities for all of the boards are also shown. In this case, the *p*-value is less than 0.05, so you reject the test of equality at the 0.05 level.

The "Predicted Values and Limits" table provides the predicted failure count and prediction limits for the number of capacitors that would fail in a future sample of size 1000 for the pooled data, as requested with the PREDICT(1000) option. The "Sample Size for Estimation" table gives the sample size required to estimate the binomial probability to within 0.05 for the pooled data, as requested with the TOLERANCE(.05) option.

The "Estimates by Group" table supplies the estimated binomial probability, confidence limits, and the contribution to the total chi-squared for each board. The pooled values are shown on the last line of the table.

The "Predicted Values by Group" table gives the predicted counts in a future sample of size 1000, prediction limits, and the sample size required to estimate the binomial probability to within the tolerance of 0.05 for each board. Values for the pooled data are shown on the last line of the table.

```
                        The RELIABILITY Procedure

                      Model Information - All Groups

                   Input Data Set              WORK.BINEX
                   Events Variable                   fail
                   Trials Variable                 sample
                   Distribution                  Binomial
                   Confidence Coefficient             95%
                   Observations Used                    9

                          Binomial Data Analysis

               Pooled Events                              81.0000
               Pooled Trials                            2058.0000
               Estimate of Proportion                      0.0394
               Lower Limit For Proportion                  0.0314
               Upper Limit For Proportion                  0.0487
               ChiSquare                                  56.8504
               Pr>ChiSquare                                0.0000

                        Predicted Value and Limits

               Sample Size For Prediction               1000.0000
               Predicted Count                            39.3586
               Lower Prediction Limit                     24.8424
               Upper Prediction Limit                     56.3237

                       Sample Size For Estimation

                 Tolerance                         0.0500
                 Sample Size For Tolerance        58.0975

                            Estimates By Group
                                        95% Confidence Limits
Group    Events    Trials         Prop        Lower        Upper           X2

1             2        84       0.0238       0.0029       0.0834       0.5371
2             3        72       0.0417       0.0087       0.1170       0.0101
3             5        72       0.0694       0.0229       0.1547       1.7237
4            19       119       0.1597       0.0990       0.2381      45.5528
5            21       538       0.0390       0.0243       0.0590       0.0015
6             2        51       0.0392       0.0048       0.1346       0.0000
7             9       517       0.0174       0.0080       0.0328       6.5884
8            18       462       0.0390       0.0233       0.0609       0.0019
9             2       143       0.0140       0.0017       0.0496       2.4348
Pooled       81      2058       0.0394       0.0314       0.0487      56.8504

                   Predicted/Tolerance Values By Group
                Predicted      95% Prediction Limits       Tolerance
      Group         Count         Lower        Upper     Sample Size

      1             23.81        1.5476      88.5824           35.71
      2             41.67        6.9416     124.6142           61.36
      3             69.44       20.4052     165.3499           99.30
      4            159.66       91.9722     254.5444          206.17
      5             39.03       20.1599      64.7140           57.64
      6             39.22        3.3970     144.2494           57.90
      7             17.41        5.3506      36.7531           26.28
      8             38.96       19.3343      66.3850           57.53
      9             13.99        0.3851      53.0715           21.19
      Pooled        39.36       24.8424      56.3237           58.10
```

Figure 30.24. Analysis of the Capacitor Data

Syntax

Primary Statements

The following are the primary statements that control the RELIABILITY procedure:

PROC RELIABILITY *<options>;*

ANALYZE *variable<∗censor-variable(values)> <=(group-variables)> </ options>;*

MCFPLOT *variable<∗cost/censor-variable(values)> <=(group-variables)> </ options>;*

MODEL *variable<∗censor-variable(values)> =<independent-variables> </ options>;*

PROBPLOT *variable<∗censor-variable(values)> <=(group-variables)> </ options>;*

RELATIONPLOT *variable<∗censor-variable(values)> <=(group-variables)> </ options>;*

The PROC RELIABILITY statement invokes the procedure.

The plot statements (PROBPLOT, RELATIONPLOT, and MCFPLOT) create graphical displays. Each of the plot statements has options that control the content and appearance of the plots they create. The default settings provide the best plots for many purposes; however, if you want to control specific details of the plots, such as axis limits or background colors, then you need to specify the options.

In addition to graphical output, each plot statement provides analysis results in tabular form. The tabular output also can be controlled with statement options.

The MODEL and ANALYZ1E statements produce only tabular analysis output, not graphical displays.

You can specify one or more of the plot and ANALYZE statements. If you specify more than one MODEL statement, only the last one specified is used.

Secondary Statements

You can specify the following statements in conjunction with the primary statements listed previously. These statements are used to modify the behavior of the primary statements or to specify additional variables.

BY *variables;*

CLASS *variables;*

DISTRIBUTION *distribution-name;*

FMODE *keyword = variable('value1' . . . 'valuen');*

FREQ *variable;*

INSET *keyword-list< options>;*

MAKE *'table'* **OUT=***SAS-data-set* < / *options*>;
NENTER *variable*;
UNITID *variable*;

The BY statement specifies variables in the input data set that are used for BY processing. A separate analysis is performed for each group of observations defined by the levels of the BY variables. The input data set must be sorted in order of the BY variables.
The CLASS statement specifies variables in the input data set that serve as *indicator, dummy,* or *classification* variables in the MODEL statement.

The DISTRIBUTION statement specifies a probability distribution name for those statements that require a probability distribution for proper operation (the ANALYZE, PROBPLOT, MODEL, and RELATIONPLOT statements). If you do not specify a distribution with the DISTRIBUTION statement, the normal distribution is used.

The FMODE statement specifies what failure-mode data to include in the analysis of data. Use this statement in conjunction with the ANALYZE, MODEL, PROBPLOT, or RELATIONPLOT statements.

The FREQ statement specifies a variable that provides frequency counts for each observation in the input data set.

The INSET statement specifies what information is printed in the inset box created by the PROBPLOT or MCFPLOT statements. The INSET statement also controls the appearance of the inset box.

The MAKE statement creates a SAS data set from any of the tables produced by the procedure. You specify a table and a SAS data set name for the data set you want to create. There is a unique table name that identifies each table printed; see the tables in the "MAKE Statement" section.

The NENTER statement specifies interval-censored data having a special structure; these data are called *readout* data. Use the NENTER statement in conjunction with the FREQ statement.

The UNITID statement specifies a variable in the input data set that is used to identify each individual unit in an MCFPLOT statement.

Graphical Enhancement Statements

You can use the TITLE, FOOTNOTE, and NOTE statements to enhance printed output. If you are creating plots, you can also use the LEGEND and SYMBOL statements to enhance your plots. For details, refer to *SAS/GRAPH Software: Reference* and the section for the plot statement that you are using.

ANALYZE Statement

ANALYZE *variable*<∗*censor-variable(values)*> <**=(***group-variables***)**>
< **/** *options* >**;**
ANALYZE *(variable1 variable2)* <**=***group-variables*> <**/** *options*>**;**

ANALYZE *variable1***(***variable2***)** <**=(***group-variables***)**> < **/** *options* >**;**

You use the ANALYZE statement to estimate the parameters of the probability distribution specified in the DISTRIBUTION statement without producing any graphical output. The ANALYZE statement performs the same analysis as the PROBPLOT statement, but it does not produce any plots. In addition, you can use the ANALYZE statement to analyze data with the binomial and Poisson distributions. The third format for the ANALYZE statement shown above applies only to Poisson and binomial data. You can use any number of ANALYZE statements after a PROC RELIABILITY statement; each ANALYZE statement produces a separate analysis.

You must specify one *variable*. If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values.

If you are using the binomial or Poisson distributions, you must specify *variable1* to represent a binomial or Poisson count and *variable2* to provide an exposure measure for the Poisson distribution or the binomial sample size for the binomial distribution.

You can optionally specify one or two *group-variables*. The ANALYZE statement produces an analysis for each level combination of the *group-variable* values. The observations in a given level are referred to as a *cell*.

The elements of the ANALYZE statement are described as follows.

variable
: represents the data for which an analysis is to be produced. A *variable* must be a numeric variable in the input data set.

censor-variable(values)
: indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you omit the specification of *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(*variable1 variable2)*
: is another method of specifying the data. You can use this syntax in a situation where uncensored, interval-censored, left-censored and right-censored values occur in the same set of data. Table 30.20 on page 973 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values. See “Lognormal Analysis with Arbitrary Censoring” on page 939 for an example of using this syntax to create a probability plot.

variable1
: represents the count data for which a Poisson or binomial analysis is to be produced. A *variable1* must be a numeric variable in the input data set.

variable2
: provides either an exposure measure for a Poisson analysis or a binomial number of trials for a binomial analysis. A *variable2* must be a numeric variable in the input data set.

group-variables
: are one or two group variables. If no group variables are specified, a single analysis is produced. The *group-variables* can be numeric or character variables in the input data set.

  Note that the parentheses surrounding the *group-variables* are needed only if two group variables are specified.

options
: control the features of the analysis. All *options* are specified after a slash (/) in the ANALYZE statement.

Summary of Options

The following tables summarize the options available in the ANALYZE statement. You can specify one or more of these options to control the parameter estimation and provide optional analyses.

Table 30.4. Analysis Options for Distributions Other than Poisson or Binomial

Option	Option Description
CONFIDENCE=*number*	specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95
CONVERGE=*number*	specifies the convergence criterion for maximum likelihood fit. See “Maximum Likelihood Estimation” on page 1004 for details.
CONVH=*number*	specifies the convergence criterion for the relative Hessian convergence criterion See “Maximum Likelihood Estimation” on page 1004 for details.
CORRB	requests parameter correlation matrix
COVB	requests parameter covariance matrix
FITTYPE | FIT=	specifies method of estimating distribution parameters
LSYX	-least squares fit to the probability plot. The probability axis is the dependent variable.
| LSXY	-least squares fit to the probability plot. The lifetime axis is the dependent variable.
| MLE	-maximum likelihood (default)
| NONE	-no fit is computed
| WEIBAYES	-Weibayes

Table 30.4. Analysis Options for Distributions Other than Poisson or Binomial (continued)

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td><(CONFIDENCE|CONF=<i>number</i>)></td><td><i>number</i> is the confidence coefficient for the Weibayes fit and is between 0 and 1. The default is 0.95.</td></tr>
<tr><td>ITPRINT</td><td>requests iteration history for maximum likelihood fit</td></tr>
<tr><td>LRCL</td><td>requests likelihood ratio confidence intervals for distribution parameters</td></tr>
<tr><td>LRCLPER</td><td>requests likelihood ratio confidence intervals for distribution percentiles</td></tr>
<tr><td>LOCATION=number < LINIT ></td><td>specifies fixed or initial value of location parameter</td></tr>
<tr><td>MAXIT=number</td><td>specifies maximum number of iterations allowed for maximum likelihood fit</td></tr>
<tr><td>MAXITEREM |
MAXITEM=<i>number1</i> <,<i>number2</i>></td><td><i>number1</i> specifies maximum number of iterations allowed for Turnbull algorithm. Iteration history will be printed in increments of <i>number2</i> if requested with ITPRINTEM. See “Interval-Censored Data” on page 999 for details.</td></tr>
<tr><td>NOPCTILES</td><td>suppress computation of percentiles</td></tr>
<tr><td>NOPOLISH</td><td>suppress setting small interval probabilities to zero in Turnbull algorithm. See “Interval-Censored Data” on page 999 for details.</td></tr>
<tr><td>PCTLIST=number-list</td><td>specifies list of percentages for which to compute percentile estimates. number-list must be a list of numbers separated by blanks or commas. Each number in the list must be between 0 and 100</td></tr>
<tr><td>PPOS=</td><td>specifies plotting position type. See “Probability Plotting” beginning on page page 995 for details.</td></tr>
<tr><td>EXPRANK</td><td>-expected ranks</td></tr>
<tr><td>|MEDRANK</td><td>-median ranks</td></tr>
<tr><td>|MEDRANK1</td><td>-median ranks (exact formula)</td></tr>
<tr><td>|KM</td><td>-Kaplan-Meier</td></tr>
<tr><td>|MKM</td><td>-modified Kaplan-Meier (default)</td></tr>
<tr><td>PPOUT</td><td>request table of cumulative probabilities</td></tr>
<tr><td>PROBLIST=number-list</td><td>specifies list of initial values for Turnbull algorithm. See “Interval-Censored Data” on page 999 for details.</td></tr>
</table>

Table 30.4. Analysis Options for Distributions Other than Poisson or Binomial (continued)

Option	**Option Description**
PSTABLE=*number*	specifies stable parameterization. The *number* must be between zero and one. See "Stable Parameters" on page 1007 for further information.
READOUT	analyze readout data
SCALE=*number* < SCINIT >	specifies fixed or initial value of scale parameter
SHAPE=*number* < SHINIT >	specifies fixed or initial value of shape parameter
SINGULAR=*number*	specifies singularity criterion for matrix inversion
SURVTIME=*number-list*	requests survival function be computed for values in *number-list*
THRESHOLD=*number*	specifies a fixed threshold parameter. See Table 30.37 for the distributions with a threshold parameter.
TOLLIKE=*number*	specifies criterion for convergence in the Turnbull algorithm. Default is 10^{-8}. See “Interval-Censored Data” on page 999 for details.
TOLPROB=*number*	specifies criterion for setting interval probability to zero in the Turnbull algorithm. Default is 10^{-6}. See “Interval-Censored Data” on page 999 for details.

Table 30.5. Analysis Options for Poisson And Binomial Distributions

Option	**Option Description**
CONFIDENCE=*number*	specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95
PREDICT(*number*)	requests predicted counts for exposure *number* for Poisson or sample size *number* for binomial
TOLERANCE(*number*)	requests exposure for Poisson or sample size for binomial to estimate Poisson rate or binomial probability within *number* with probability given by the CONFIDENCE= option

CLASS Statement

CLASS *variable-names*

The CLASS statement specifies variables in the input data set that serve as *indicator*, *dummy*, or *classification* variables in the MODEL statement. If a CLASS variable is specified as an independent variable in the MODEL statement, the RELIABILITY procedure automatically generates an indicator variable for each level of the CLASS variable. The indicator variables generated are used as independent variables in the regression model specified in the MODEL statement. An indicator variable for a level of a CLASS variable is a variable equal to 1 for those observations corresponding to the level and equal to 0 for all other observations.

DISTRIBUTION Statement

DISTRIBUTION *probability distribution-name*

The ANALYZE, PROBPLOT, RELATIONPLOT, and MODEL statements require you to specify the probability distribution that describes your data. You can specify a probability distribution using the DISTRIBUTION statement anywhere after the PROC RELIABILITY statement and before the RUN statement. If you do not specify a distribution in a DISTRIBUTION statement, the normal distribution is assumed. The probability distribution specified determines the distribution for which parameters are estimated using your data. The valid distributions and the statements to which they apply are shown in Table 30.6.

Table 30.6. Probability Distributions

Distribution	Distribution-Name Specified	Statement
binomial	BINOMIAL	ANALYZE
exponential	EXPONENTIAL	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
extreme value	EXTREME | EV	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
generalized gamma	GAMMA	MODEL
logistic	LOGISTIC | LOGIT	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
loglogistic	LLOGISTIC | LLOGIT	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
lognormal (base e)	LOGNORMAL | LNORM	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
lognormal (base 10)	LOGNORMAL10 | LNORM10	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
normal	NORMAL	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Poisson	POISSON	ANALYZE
Weibull	WEIBULL	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL

FMODE Statement

FMODE *keyword***=** *variable ('value1' . . . 'valuen')***;**

Use the FMODE statement with data that have failures attributable to multiple causes, or *failure modes*. You can analyze data by either keeping or eliminating specific failure modes with the FMODE statement. Use this statement in conjunction with the ANALYZE, MODEL, PROBPLOT, or RELATIONPLOT statements. An FMODE statement placed anywhere after the PROC RELIABILITY statement and before the RUN statement affects analyses performed by all ANALYZE, MODEL, PROBPLOT, and RELATIONPLOT statements.

If you specify the keyword KEEP, the life distribution for only the identified failure modes is estimated, with all other failure modes treated as right-censored data. If you specify the keyword ELIMINATE, the life distribution that results if the failure modes identified are completely eliminated is estimated. The keyword ELIMINATE causes the failure modes identified to be treated as right-censored data and causes a single life distribution to be estimated for the remaining data. The failure mode for an observation in the input data set is identified by the value of *variable*, where *variable* is any numeric or character variable in the input data set. You specify failure modes to keep or eliminate by listing variable-values *('value1' . . . 'valuen')* in parentheses after the failure mode variable name. The list of variable-values must have each entry enclosed in single quotes, with entries separated by blanks or commas.

FREQ Statement

FREQ *variable-name*

The FREQ statement specifies a variable that provides frequency counts for each observation in the input data set. If n is the value of the FREQ variable for an observation, then that observation is weighted by n. The log-likelihood function for maximum likelihood estimation is multiplied by n. If n is not an integer, the integer part of n is used in creating probability plots.

You can also use the FREQ statement in conjunction with the NENTER statement to specify interval-censored data having a special structure; these data are called *readout* data. The FREQ statement specifies a variable in the input data set that determines the number of units failing in each interval. See "Weibull Analysis of Interval Data with Common Inspection Schedule" on page 935 for an example using the FREQ statement with readout data.

INSET Statement

INSET *keyword-list*$<$ **/** *options* $>$**;**

The box or table of summary information produced on plots made with the PROBPLOT or MCFPLOT statement is called an *inset*. You can use the INSET statement to customize the information that is printed in the inset box and the appearance of the inset box. To supply the information that is displayed in the inset box, you specify *keywords* corresponding to the information you want shown. For example, the following statements produce a Weibull plot with the sample size, the number of failures, and the Weibull mean displayed in the inset.

```
proc reliability data=fan;
   distribution weibull;
   pplot lifetime*censor(1);
   inset n nfail weibull(mean);
run;
```

By default, inset entries are identified with appropriate labels. However, you can provide a customized label by specifying the *keyword* for that entry followed by the equal sign (=) and the label in quotes. For example, the following INSET statement produces an inset containing the sample size and Weibull mean, labeled "Sample Size" and "Weibull Mean" in the inset.

```
inset n='Sample Size' weibull(mean='Weibull Mean');
```

If you specify a keyword that does not apply to the plot you are creating, then the keyword is ignored.

The *options* control the appearance of the box.

If you specify more than one INSET statement, only the last one is used.

Keywords Used in the INSET Statement

The following tables list keywords available in the INSET statement to display summary statistics, distribution parameters, and distribution fitting information.

Table 30.7. Summary Statistics

keyword	*description*
N	sample size
NFAIL	number of failures for probability plots
NEVENTS	number of events or repairs for MCF plots
NEVENTS1	number of events or repairs in the first group for MCF difference plots
NEVENTS2	number of events or repairs in the second group for MCF difference plots
NUNITS	number of units or systems for MCF plots
NUNITS1	number of units or systems in the first group for MCF difference plots
NUNITS2	number of units or systems in the second group for MCF difference plots

Table 30.8. General Information

keyword	*description*
CONFIDENCE	confidence coefficient for all confidence intervals or for the Weibayes fit
FIT	method used to estimate distribution parameters for probability plots
RSQUARE	R^2 for least squares distribution fit to probability plots

Distribution parameters are specified as *distribution-name*(*distribution-parameters*). The following table lists the keywords available.

Table 30.9. Distribution Parameters

keyword	*secondary keyword*	*description*
EXPONENTIAL	SCALE	scale parameter
	THRESHOLD	threshold parameter
	MEAN	expected value
EXTREME | EV	LOCATION	location parameter
	SCALE	scale parameter
	MEAN	expected value
LOGISTIC | LOGIT	LOCATION	location parameter
	SCALE	scale parameter
	MEAN	expected value
LOGLOGISTIC | LLOGIT	LOCATION	location parameter
	SCALE	scale parameter
	THRESHOLD	threshold parameter
	MEAN	expected value
LOGNORMAL	LOCATION	location parameter
	SCALE	scale parameter
	THRESHOLD	threshold parameter
	MEAN	expected value
LOGNORMAL10	LOCATION	location parameter
	SCALE	scale parameter
	THRESHOLD	threshold parameter
	MEAN	expected value
NORMAL	LOCATION	location parameter
	SCALE	scale parameter
	MEAN	expected value
WEIBULL	SCALE	scale parameter
	SHAPE	shape parameter
	THRESHOLD	threshold parameter
	MEAN	expected value

Options Used in the INSET Statement

The following tables list INSET statement options that control the appearance of the inset box.

Table 30.10. General Appearance Options

Option	Option Description
HEADER=*'quoted string'*	specifies text for header or box title
NOFRAME	omits frame around box
POS= *value* <DATA | PERCENT>	determines the position of the inset. The *value* can be a compass point (N, NE, E, SE, S, SW, W, NW) or a pair of coordinates (x,y) enclosed in parentheses. The coordinates can be specified in axis percent units or axis data units.
REFPOINT= *name*	specifies the reference point for an inset that is positioned by a pair of coordinates with the POS= option. You use the REFPOINT= option in conjunction with the POS= coordinates. The REFPOINT= option specifies which corner of the inset frame you have specified with coordinates (x,y) and it can take the value of BR (bottom right), BL (bottom left), TR (top right), or TL (top left). The default is REFPOINT=BL. If the inset position is specified as a compass point, then the REFPOINT= option is ignored.

Table 30.11. Text Enhancement Options

Option	Option Description
FONT=*font*	software font for text
HEIGHT=*value*	height of text

Table 30.12. Color and Pattern Options

Option	Option Description
CFILL=*color*	color for filling box
CFILLH=*color*	color for filling box header
CFRAME=*color*	color for frame
CHEADER=*color*	color for text in header
CTEXT=*color*	color for text

MAKE Statement

MAKE *'table'* **OUT=***SAS-data-set*<*(SAS-data-set options)*>**;**

The MAKE statement creates a SAS data set from any of the tables produced by the RELIABILITY procedure. You can specify SAS data set options in parentheses after the data set name. You can specify one MAKE statement for each table that you want to save to a SAS data set.

The ODS statement also creates SAS data sets from tables, in addition to providing an extensive and flexible method of controlling output created by the RELIABILITY procedure. The ODS statement is the recommended method of controlling procedure output, however, the MAKE statement is provided for compatibility with earlier releases of the SAS system

The valid values for *table* are shown in "ODS Table Names" on page 1025, organized by the RELIABILITY procedure statement that produces the tabular output. The *table* names are not case sensitive, but they must be enclosed in single quotes.

MCFPLOT Statement

MCFPLOT *variable* ∗*cost/censor-variable(values)* <**=(***group-variables***)**> <**/***options*>**;**

You can specify any number of MCFPLOT statements after a PROC RELIABILITY statement. Each MCFPLOT statement creates a separate MCF plot and associated analysis. See "Analysis of Recurrence Data on Repairs" on page 947 and "Comparison of Two Samples of Repair Data" on page 949 for examples using the MCFPLOT statement.

To create a mean cumulative function plot for cost or number of repairs, you specify a *variable* that represents the times of repairs. You must also specify a *cost/censor-variable* and the *values*, in parentheses, of the *cost/censor-variable* that correspond to end-of-history data values (also referred to as *censored* data values).

You can optionally specify one or two *group-variables* (also referred to as *classification variables*). The MCFPLOT statement displays a component plot for each level of the *group-variables* using the values of the *variable*. The observations in a given level are referred to as a *cell*.

You must also specify a *unit-identification* variable in conjunction with the MCFPLOT statement to identify the individual unit name for each instance of repair or end of history on the unit. Specify the *unit-identification* variable in the UNITID statement.

The elements of the MCFPLOT statement are described as follows.

variable

represents the time of repair. A *variable* must be a numeric variable in the input data set.

cost/censor-variable(values)

indicates the cost of each repair or the number of repairs. This variable also indicates which observations in the input data set are end-of-history (censored) data points. You specify the values of *cost/censor-variable* that represent censored observations by placing those values in parentheses after the variable name. A *censor-variable* must be a numeric variable in the input data set.

group-variables

are one or two group variables. If no group variables are specified, a single plot is produced. The *group-variables* can be any numeric or character variables in the input data set.

Note that the parentheses surrounding the *group-variables* are needed only if two group variables are specified.

options

control the features of the mean cumulative function plot. All *options* are specified after a slash (/) in the MCFPLOT statement. The "Summary of Options" section, which follows, lists all options by function.

Summary of Options

Table 30.13. Analysis Options

Option	**Option Description**
CONFIDENCE=*number*	specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95
INDINC	requests variance estimates of the MCF computed under an independent increments assumption
MCFDIFF	requests a plot of differences of MCFS of two groups specified by a single group variable
NOVARIANCE	suppresses MCF variance computation
VARMETHOD2	requests the method of Lawless and Nadeau (1995) be used to compute variance estimates of the MCF

Table 30.14. Plot Layout Options

Option	Option Description
CENBIN	plots censored data as frequency counts rather than as individual points
CENSYMBOL=*symbol* | (*symbol list*)	specifies symbols for censored values. *symbol* is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z). If you are creating overlaid plots for groups of data, you can specify different symbols for the groups with a list of symbols or letters, separated by blanks, enclosed in parentheses. If no CENSYMBOL option is specified, the symbol used for censored values is the same as for repairs.
HOFFSET=*value*	specifies offset for horizontal axis
INBORDER	requests a border around MCF plots
INTERTILE=*value*	specifies distance between tiles
JITTER=*number*	specifies amount to jitter overlaying plot symbols, in units of symbol width
MCFLEGEND=*legend-statement-name* | NONE	identifies legend statement to specify legend for overlaid MCF plots
MISSING1	requests that missing values of first GROUP= variable be treated as a level of the variable
MISSING2	requests that missing values of second GROUP= variable be treated as a level of the variable
NCOLS=n	specifies number of columns plotted on a page
NOCENPLOT	suppresses plotting of censored data points
NOCONF	suppresses plotting of confidence intervals
NOFRAME	suppresses frame around plotting area
NOINSET	suppresses inset
NOLEGEND	suppresses legend for overlaid MCF plots
NROWS=n	specifies number of rows plotted on a page
ORDER1=DATA | FORMATTED | FREQ | INTERNAL	specifies display order for values of the first GROUP= variable
ORDER2=DATA | FORMATTED | FREQ | INTERNAL	specifies display order for values of the second GROUP= variable
OVERLAY	requests plots with group variables be overlaid on a single page
TURNVLABELS	vertically strings out characters in labels for vertical axis
VOFFSET=*value*	specifies length of offset at upper end of vertical axis

Table 30.15. Reference Line Options

Option	**Option Description**
HREF=*value-list*	specifies reference lines perpendicular to horizontal axis
HREFLABELS=*('label1' . . . 'labeln')*	specifies labels for HREF= lines.
HREFLABPOS=*n*	specifies vertical position of labels for HREF= lines. The valid values for *n* and the corresponding label placements are shown below. *n* | **label placement** 1 | top 2 | staggered from top 3 | bottom 4 | staggered from bottom 5 | alternating from top 6 | alternating from bottom
LHREF=*linetype*	specifies line style for HREF= lines
LVREF=*linetype*	specifies line style for VREF= lines
VREF=*value-list*	specifies reference lines perpendicular to vertical axis
VREFLABELS=*('label1' . . . 'labeln')*	specifies labels for VREF= lines
VREFLABPOS=*n*	specifies horizontal position of labels for VREF= lines. The valid values for *n* and the corresponding label placements are shown below. *n* | **label placement** 1 | left 2 | right

Table 30.16. Text Enhancement Options

Option	**Option Description**
FONT=*font*	software font for text
HEIGHT=*value*	height of text used outside framed areas
INFONT=*font*	software font for text inside framed areas
INHEIGHT=*value*	height of text inside framed areas

Table 30.17. Axis Options

Option	Option Description

HAXIS=*value1 to value2<by value3>*	specifies tick mark values for the horizontal axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of HAXIS= lists follow:
	`haxis = 0 to 10`
	`haxis = 2 to 10 by 2`
	`haxis = 0 to 200 by 10`

HLOWER=*number*	specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies *number* as the lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the HAXIS option is used.
HUPPER=*number*	specifies the upper limit on the horizontal axis scale. The HUPPER= option specifies *number* as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the HAXIS= option is used.
LGRID=*number*	specifies a line style for all grid lines. *number* is between 1 and 46 and specifies a linestyle for grids.
LOGLOG	requests log scales on both axes
MINORLOGGRID	adds a minor grid for log axes
NOGRID	suppresses grid lines
NOHLABEL	suppresses label for horizontal axis
NOVLABEL	suppresses label for vertical axis
NOVTICK	suppresses tick marks and tick mark labels for vertical axis
NOHTICK	suppresses tick marks and tick mark labels for horizontal axis
NHTICK=*number*	specifies number of tick intervals for the horizontal axis. This option has no effect if the HAXIS= option is used.

Table 30.17. Axis Options (continued)

Option	**Option Description**
NVTICK=*number*	specifies number of tick intervals for the vertical axis. This option has no effect if the VAXIS= option is used.

VAXIS=*value1 to value2<by value3>*	specifies tick mark values for the vertical axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. This method of specification of tick marks is not valid for logarithmic axes. If *value3* is omitted, a value of 1 is used.
	`vaxis = 0 to 10`
	`vaxis = 0 to 2 by .1`

VAXISLABEL='*string*'	specifies a label for the vertical axis
VLOWER=*number*	specifies the lower limit on the vertical axis scale. The VLOWER= option specifies *number* as the lower vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the VAXIS= option is used.
VUPPER=*number*	specifies the upper limit on the vertical axis scale. The VUPPER= option specifies *number* as the upper vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the VAXIS= option is used.
WAXIS=*n*	specifies line thickness for axes and frame

Table 30.18. Color and Pattern Options

Option	**Option Description**
CAXIS=*color*	color for axis
CCENSOR=*color*	color for filling censor plot area
CENCOLOR=*color*	color for censor symbol
CFRAME=*color*	color for frame
CFRAMESIDE=*color*	color for filling frame for row labels
CFRAMETOP=*color*	color for filling frame for column labels
CGRID=*color*	color for grid lines
CHREF=*color*	color for HREF= lines
CTEXT=*color*	color for text
CVREF=*color*	color for VREF= lines

Table 30.19. Graphics Catalog Options

Option	Option Description
DESCRIPTION=‘*string*’	description for graphics catalog member
NAME=‘*string*’	name for plot in graphics catalog

MODEL Statement

MODEL *variable*<∗*censor-variable(values)*> <**=***effect-list*> < **/** *options* >**;**
MODEL *(variable1 variable2)* <**=***effect-list*> <**/***options*>**;**

You use the MODEL statement to fit regression models, where life is modeled as a function of explanatory variables.

You can use only one MODEL statement after a PROC RELIABILITY statement. If you specify more than one MODEL statement, only the last is used.

The MODEL statement does not produce any plots, but it enables you to analyze more complicated regression models than the ANALYZE, PROBPLOT, or RELATIONPLOT statement does. The probability distribution specified in the DISTRIBUTION statement is used in the analysis. The following are examples of MODEL statements:

```
model time = temp voltage;
model life*censor(1) = voltage width;
```

See “Analysis of Accelerated Life Test Data” on page 930 and “Regression Modeling” on page 943 for examples of fitting regression models using the MODEL statement.

If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values.

If your data contain any interval-censored or left-censored values, you must specify *variable1* and *variable2* in parentheses to provide the endpoints of the interval for each observation.

The independent variables in your regression model are specified in the *effect-list*. The *effect-list* is any combination of continuous variables, classification variables,

See Regression Models on page 1006 for further information on specifying the independent variables.

The elements of the MODEL statement are described as follows.

variable
is the dependent, or response, variable. The *variable* must be a numeric variable in the input data set.

censor-variable(values)
indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those

values in parentheses after the variable name. If your data are not right censored, then you can omit the specification of a *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(*variable1 variable2)*

is another method of specifying the dependent variable in the regession model. You can use this syntax in a situation where uncensored, interval-censored, left-censored and right-censored values occur in the same set of data. Table 30.20 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values.

Table 30.20. Specifying Censored Values

Variable1	Variable2	Type of Censoring
nonmissing	nonmissing	uncensored if *variable1* = *variable2*
nonmissing	nonmissing	interval censored if *variable1* < *variable2*
nonmissing	missing	right censored at *variable1*
missing	nonmissing	left censored at *variable2*

For example, if T1 and T2 represent time in hours in the input data set

```
OBS    T1    T2

 1      .     6
 2      6    12
 3     12    24
 4     24     .
 5     24    24
```

then the statement

```
model (t1 t2);
```

specifies a model in which observation 1 is left censored at 6 hours, observation 2 is interval censored in the interval (6, 12), observation 3 is interval censored in (12,24), observation 4 is right censored at 24 hours, and observation 5 is an uncensored lifetime of 24 hours.

effect-list

is a list of variables in the input data set representing the values of the independent variables in the model for each observation, and combinations of variables representing interaction terms. If a variable in the *effect-list* is also listed in a CLASS statement, an indicator variable is generated for each level of the variable. An indicator variable for a particular level is equal to 1 for observations with that level, and equal to 0 for all other observations. This type of variable is called a *classification* variable. Classification variables can be either character or numeric. If a variable is not listed in a CLASS statement, it is assumed to be a continuous variable, and it must be numeric.

options
control how the model is fit and what output is produced. All *options* are specified after a slash (/) in the MODEL statement. The “Summary of Options” section, which follows, lists all options by function.

Summary of Options

Table 30.21. Model Statement Options

Option	Option Description
CONFIDENCE=*number*	specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95
CONVERGE=*number*	specifies the convergence criterion for maximum likelihood fit. See “Maximum Likelihood Estimation” on page 1004 for details.
CONVH=*number*	specifies the convergence criterion for the relative Hessian convergence criterion See “Maximum Likelihood Estimation” on page 1004 for details.
CORRB	requests parameter correlation matrix
COVB	requests parameter covariance matrix
INITIAL=*number list*	specifies initial values for regression parameters other than the location, or intercept term
ITPRINT	requests iteration history for maximum likelihood fit
LRCL	requests likelihood ratio confidence intervals for distribution parameters
LOCATION=*number* < LINIT >	specifies fixed or initial value of the location, or intercept parameter
MAXIT=*number*	specifies maximum number of iterations allowed for maximum likelihood fit
OBSTATS	requests a table containing the XBETA, SURV, SRESID, and ADJRESID statistics in Table 30.22. The table also contains the dependent and independent variables in the model.
OBSTATS(*statistics*)	requests a table containing the model variables and the statistics in the specified list of *statistics*. Available statistics are shown in Table 30.22.
ORDER=DATA | FORMATTED | FREQ | INTERNAL	specifies sort order for values of the classification variables in the *effect-list*
PSTABLE=*number*	specifies stable parameterization. The *number* must be between zero and one. See "Stable Parameters" on page 1007 for further information.

Table 30.21. Model Statement Options (continued)

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>READOUT</td><td>analyzes data in readout structure. The FREQ statement must be used to specify the number of units failing in each interval, and the NENTER statement must be used to specify the number of unfailed units entering each interval</td></tr>
<tr><td>RELATION=ARRHENIUS |
ARRHENIUS2 |
POWER
RELATION=(ARRHENIUS |
ARRHENIUS2 |
POWER
< , >
ARRHENIUS |
ARRHENIUS2 |
POWER)</td><td>specifies type of relationship between independent and dependent variables. In the first form, the transformation specified is applied to the first continuous independent variable in the model. In the second form, the transformations specified within parentheses are applied to the first two continuous independent variables in the model, in the order listed. See Table 30.45 on page 1007 for definitions of the transformations.</td></tr>
<tr><td>SCALE=number < SCINIT ></td><td>specifies fixed or initial value of scale parameter</td></tr>
<tr><td>SHAPE=number < SHINIT ></td><td>specifies fixed or initial value of shape parameter</td></tr>
<tr><td>SINGULAR=number</td><td>specifies singularity criterion for matrix inversion</td></tr>
<tr><td>THRESHOLD=number</td><td>specifies a fixed threshold parameter. See Table 30.37 for the distributions with a threshold parameter.</td></tr>
</table>

Table 30.22. Observation Statistics Available in the OBSTATS Option

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>CENSOR</td><td>is an indicator variable equal to 1 if an observation is censored, and 0 otherwise</td></tr>
<tr><td>QUANTILES | QUANTILE |
Q=number list</td><td>specifies distribution quantiles for each number in number list for each observation. The numbers must be between 0 and 1. Estimated quantile standard errors, and upper and lower confidence limits are also tabulated.</td></tr>
<tr><td>XBETA</td><td>is the linear predictor</td></tr>
<tr><td>SURVIVAL | SURV</td><td>is the fitted survival function, evaluated at the value of the dependent variable</td></tr>
<tr><td>RESID</td><td>is the raw residual</td></tr>
<tr><td>SRESID</td><td>is the standardized residual</td></tr>
<tr><td>GRESID</td><td>is the modified Cox-Snell residual</td></tr>
<tr><td>DRESID</td><td>is the deviance residual</td></tr>
</table>

Table 30.22. Observation Statistics Available in the OBSTATS Option (continued)

Option	**Option Description**
ADJRESID	is the adjusted standardized residuals. These are adjusted for right-censored observations by adding the median of the lifetime above the right-censored values to the residuals.
RESIDADJ=*number*	specifies adjustment to be added to Cox-Snell residual for right-censored data values. The default is log(2) = 0.693.
RESIDALPHA | RALPHA=*number*	specifies $number \times 100\%$ percentile residual lifetime used to adjust right-censored standardized residuals. The *number* must be between 0 and 1. The default value is 0.5, corresponding to the median.
CONTROL=*variable*	specifies a control variable in the input data set. If the value of the control variable is 1, the observation statistics are computed. If the value of the control variable is not equal to 1, the statistics are not computed for that observation.

NENTER Statement

NENTER *variable*

Use the NENTER statement in conjunction with the FREQ statement to specify interval-censored data having a special structure; these data are called *readout* data. The NENTER statement specifies a *variable* in the input data set that determines the number of unfailed units entering each interval. See “Weibull Analysis of Interval Data with Common Inspection Schedule” on page 935 for an example using the NENTER statement with readout data.

PROBPLOT Statement

PROBPLOT *variable*<∗*censor-variable(values)*> <**=(***group-variables***)**> <**/***options*>**;**
PROBPLOT *(variable1 variable2)* <**=***group-variables*> <**/***options*>**;**

You use the PROBPLOT statement to create a probability plot from complete, left-censored, right-censored, or interval censored data.

You can specify the keyword PPLOT as an alias for PROBPLOT. You can specify any number of PROBPLOT statements after a PROC RELIABILITY statement. Each PROBPLOT statement creates a probability plot and an associated analysis. The probability distribution used in creating the probability plot and performing the analysis is determined by the DISTRIBUTION statement.

See “Analysis of Right-Censored Data from a Single Population” on page 925 and “Weibull Analysis Comparing Groups of Data” on page 928 for examples creating probability plots using the PROBPLOT statement.

To create a probability plot, you must specify one *variable*. If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values.

You can optionally specify one or two *group-variables* (also referred to as *classification variables*). The PROBPLOT statement displays a component probability plot for each level of the *group-variables* using the values of the *variable*. The observations in a given level are referred to as a *cell*.

The elements of the PROBPLOT statement are described as follows.

variable
: represents the data for which a probability plot is to be produced. The *variable* must be a numeric variable in the input data set.

censor-variable(values)
: indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you can omit the specification of *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(*variable1 variable2)*
: is another method of specifying the data for which a probability plot is to be produced. You can use this syntax in a situation where uncensored, interval-censored, left-censored and right-censored values occur in the same set of data. Table 30.20 on page 973 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values. See “Lognormal Analysis with Arbitrary Censoring” on page 939 for an example of using this syntax to create a probability plot.

group-variables
: are one or two group variables. If no group variables are specified, a single probability plot is produced. The *group-variables* can be numeric or character variables in the input data set.

: Note that the parentheses surrounding the *group-variables* are needed only if two group variables are specified.

options
: control the features of the probability plot. All *options* are specified after the slash (/) in the PROBPLOT statement. The “Summary of Options” section on page 978, which follows, lists all options by function.

Summary of Options

Table 30.23. Analysis Options

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>CONFIDENCE=number</td><td>specifies the confidence coefficient for all confidence intervals. The number must be between 0 and 1. The default value is 0.95</td></tr>
<tr><td>CONVERGE=number</td><td>specifies the convergence criterion for maximum likelihood fit. See “Maximum Likelihood Estimation” on page 1004 for details.</td></tr>
<tr><td>CONVH=number</td><td>specifies the convergence criterion for the relative Hessian convergence criterion See “Maximum Likelihood Estimation” on page 1004 for details.</td></tr>
<tr><td>CORRB</td><td>requests parameter correlation matrix</td></tr>
<tr><td>COVB</td><td>requests parameter covariance matrix</td></tr>
<tr><td>FITTYPE | FIT=</td><td>specifies method of estimating distribution parameters</td></tr>
<tr><td>LSYX</td><td>-least squares fit to the probability plot. The probability axis is the dependent variable.</td></tr>
<tr><td>| LSXY</td><td>-least squares fit to the probability plot. The lifetime axis is the dependent variable.</td></tr>
<tr><td>| MLE</td><td>-maximum likelihood (default)</td></tr>
<tr><td>| MODEL</td><td>-use the fit from the preceding MODEL statement</td></tr>
<tr><td>| NONE</td><td>-no fit is computed</td></tr>
<tr><td>| WEIBAYES <(CONFIDENCE| CONF=<i>number</i>)></td><td>-Weibayes method
<i>number</i> is the confidence coefficient for the Weibayes fit and is between 0 and 1. The default value is 0.95.</td></tr>
<tr><td>ITPRINT</td><td>requests iteration history for maximum likelihood fit</td></tr>
<tr><td>LRCL</td><td>requests likelihood ratio confidence intervals for distribution parameters</td></tr>
<tr><td>LRCLPER</td><td>requests likelihood ratio confidence intervals for distribution percentiles</td></tr>
<tr><td>LOCATION=number < LINIT ></td><td>specifies fixed or initial value of location parameter</td></tr>
<tr><td>MAXIT=number</td><td>specifies maximum number of iterations allowed for maximum likelihood fit</td></tr>
<tr><td>MAXITEM=<i>number1</i> <,<i>number2</i>></td><td><i>number1</i> specifies maximum number of iterations allowed for Turnbull algorithm. Iteration history will be printed in increments of <i>number2</i> if requested with ITPRINTEM. See “Interval-Censored Data” on page 999 for details.</td></tr>
<tr><td>NOPCTILES</td><td>suppresses computation of percentiles for standard list of percentage points</td></tr>
<tr><td>NOPOLISH</td><td>suppress setting small interval probabilities to zero in Turnbull algorithm. See “Interval-Censored Data” on page 999 for details.</td></tr>
</table>

Table 30.23. Analysis Options (continued)

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>PCTLIST=number-list</td><td>specifies list of percentages for which to compute percentile estimates. The number-list must be a list of numbers separated by blanks or commas. Each number in the list must be between 0 and 100.</td></tr>
<tr><td>PPOS=</td><td>specifies plotting position type. See “Probability Plotting” beginning on page 995 for details.</td></tr>
<tr><td>EXPRANK</td><td>-expected ranks</td></tr>
<tr><td>|MEDRANK</td><td>-median ranks</td></tr>
<tr><td>|MEDRANK1</td><td>-median ranks (exact formula)</td></tr>
<tr><td>|KM</td><td>-Kaplan-Meier</td></tr>
<tr><td>|MKM</td><td>-modified Kaplan-Meier (default)</td></tr>
<tr><td>PPOUT</td><td>request table of cumulative probabilities</td></tr>
<tr><td>PROBLIST=number-list</td><td>specifies list of initial values for Turnbull algorithm. See “Interval-Censored Data” on page 999 for details.</td></tr>
<tr><td>PSTABLE=number</td><td>specifies stable parameterization. The number must be between zero and one. See "Stable Parameters" on page 1007 for further information.</td></tr>
<tr><td>READOUT</td><td>analyzes data with readout structure</td></tr>
<tr><td>SCALE=number < SCINIT ></td><td>specifies fixed or initial value of scale parameter</td></tr>
<tr><td>SHAPE=number < SHINIT ></td><td>specifies fixed or initial value of shape parameter</td></tr>
<tr><td>SINGULAR=number</td><td>specifies singularity criterion for matrix inversion</td></tr>
<tr><td>SURVTIME=number-list</td><td>requests survival function for values in number-list</td></tr>
<tr><td>THRESHOLD=number</td><td>specifies a fixed threshold parameter. See Table 30.37 for the distributions with a threshold parameter.</td></tr>
<tr><td>TOLLIKE=number</td><td>specifies criterion for convergence in the Turnbull algorithm. Default is 10^{-8}. See “Interval-Censored Data” on page 999 for details.</td></tr>
<tr><td>TOLPROB=number</td><td>specifies criterion for setting interval probability to zero in the Turnbull algorithm. Default is 10^{-6}. See “Interval-Censored Data” on page 999 for details.</td></tr>
</table>

Table 30.24. Probability Plot Layout Options

Option	**Option Description**
CENBIN	plots censored data as frequency counts rather than as individual points
CENSYMBOL= *symbol* | (*symbol list*)	specifies symbols for censored values. The *symbol* is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z). For overlaid plots for groups of data, you can specify different symbols for the groups with a list of symbols or letters, separated by blanks, enclosed in parentheses. If no CENSYMBOL option is specified, the symbol used for censored values is the same as for failures.
HOFFSET=*value*	specifies offset for horizontal axis
INBORDER	requests a border around probability plots
INTERTILE=*value*	specifies distance between tiles
JITTER=*number*	specifies amount to jitter overlaying plot symbols, in units of symbol width
LFIT=*linetype*	specifies a line style for fit line and confidence curves
MISSING1	requests that missing values of first GROUP= variable be treated as a level of the variable
MISSING2	requests that missing values of second GROUP= variable be treated as a level of the variable
NCOLS=n	specifies number of columns plotted on a page
NOCENPLOT	suppresses plotting of censored data points
NOCONF	suppresses plotting of percentile confidence curves
NOFIT	suppresses plotting of fit line and percentile confidence curves
NOFRAME	suppresses frame around plotting area
NOINSET	suppresses inset
NOPPLEGEND	suppresses legend for overlaid probability plots
NROWS=n	specifies number of rows plotted on a page
ORDER1=DATA | FORMATTED | FREQ | INTERNAL	specifies display order for values of the first GROUP= variable
ORDER2=DATA | FORMATTED | FREQ | INTERNAL	specifies display order for values of the second GROUP= variable
OVERLAY	requests overlaid plots for group variables
PCONFPLT	plots confidence intervals on probabilities for readout data
PPLEGEND = *legend-statement-name* | NONE	identifies LEGEND*n* statement to specify legend for overlaid probability plots
ROTATE	requests probability plots with probability scale on horizontal axis

Table 30.24. Probability Plot Layout Options (continued)

Option	**Option Description**
TURNVLABELS	vertically strings out characters in labels for vertical axis
VOFFSET=*value*	length of offset at upper end of vertical axis
WFIT=*linetype*	line width for fit line and confidence curves

Table 30.25. Reference Line Options

Option	**Option Description**
HREF < (INTERSECT) >=*value-list*	requests reference lines perpendicular to horizontal axis. If (INTERSECT) is specified, a second reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the intersecting vertical axis reference line is labeled with the vertical axis value.
HREFLABELS=*('label1' . . . 'labeln')*	specifies labels for HREF= lines.
HREFLABPOS=*n*	specifies vertical position of labels for HREF= lines. The valid values for *n* and the corresponding label placements are shown below. *n* | **label placement** 1 | top 2 | staggered from top 3 | bottom 4 | staggered from bottom 5 | alternating from top 6 | alternating from bottom
LHREF=*linetype*	specifies line style for HREF= lines
LVREF=*linetype*	specifies line style for VREF= lines
VREF < (INTERSECT) >=*value-list*	specifies reference lines perpendicular to vertical axis. If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value.
VREFLABELS=*('label1' . . . 'labeln')*	specifies labels for VREF= lines
VREFLABPOS=*n*	specifies horizontal position of labels for VREF= lines. The valid values for *n* and the corresponding label placements are shown below. *n* | **label placement** 1 | left 2 | right

Table 30.26. Text Enhancement Options

Option	Option Description
FONT=*font*	specifies a software font for text
HEIGHT=*value*	specifies height of text used outside framed areas
INFONT=*font*	specifies a software font for text inside framed areas
INHEIGHT=*value*	specifies height of text inside framed areas

Table 30.27. Axis Options

Option	Option Description

LAXIS=*value1 to value2<by value3>*	specifies tick mark values for the lifetime axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of LAXIS= lists are
	`laxis = -1 to 10`
	`laxis = 0 to 200 by 10`

LGRID=*number*	specifies a line style for all grid lines. The *number* is between 1 and 46 and specifies a linestyle for grids.
LIFELOWER | LLOWER=*number*	specifies the lower limit on the lifetime axis scale. The LLOWER option specifies *number* as the lower lifetime axis tick mark. The tick interval and the upper lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
LIFEUPPER | LUPPER=*number*	specifies the upper limit on the lifetime axis scale. The LUPPER option specifies *number* as the upper lifetime axis tick mark. The tick interval and the lower lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
MPGRID	adds a minor grid for the probability axis
MINORLOGGRID	adds a minor grid for log axes
NOGRID	suppresses grid lines
NOLLABEL	suppresses label for life, or analysis variable, axis

Table 30.27. Axis Options (continued)

Option	Option Description
NOLTICK	suppresses tick marks and tick mark labels for lifetime or analysis variable axis
NOPLABEL	suppresses label for probability axis
NOPTICK	suppresses tick marks and tick mark labels for the probability axis
NTICK=*number*	specifies the number of tick intervals for the lifetime axis. This option has no effect if the LAXIS option is used.
PCTLOWER | PLOWER=*number*	specifies the lower limit on probability axis scale
PCTUPPER | PUPPER=*number*	specifies the upper limit on probability axis scale
PAXISLABEL='*string*'	specifies a label for the probability axis
WAXIS=n	specifies the line thickness for axes and frame

Table 30.28. Color and Pattern Options

Option	Option Description
CAXIS=*color*	color for axis
CCENSOR=*color*	color for filling censor plot area
CENCOLOR=*color*	color for censor symbol
CFIT=*color*	color for fit line and confidence curves
CFRAME=*color*	color for frame
CFRAMESIDE=*color*	color for filling frame for row labels
CFRAMETOP=*color*	color for filling frame for column labels
CGRID=*color*	color for grid lines
CHREF=*color*	color for HREF= lines
CTEXT=*color*	color for text
CVREF=*color*	color for VREF= lines

Table 30.29. Graphics Catalog Options

Option	Option Description
DESCRIPTION='*string*'	description for graphics catalog member
NAME='*string*'	name for plot in graphics catalog

RELATIONPLOT Statement

RELATIONPLOT *variable*<∗*censor-variable(values)*> <**=***group-variable*>
<**/***options*>**;**
RELATIONPLOT *(variable1 variable2)* <**=***group-variables*> <**/***options*>**;**

You use the RELATIONPLOT statement to create life-stress relation plots. A life-stress relation plot is a graphical tool for the analysis of data from accelerated life tests. The plot is a display of the relationship between life and *stress*, such as temperature or voltage. You can also use the RELATIONPLOT statement to display a probability plot alongside the relation plot. See Figure 30.6 on page 933 for an example of a relation plot.

You can specify the keyword RPLOT as an alias for RELATIONPLOT. You can use any number of RELATIONPLOT statements after a PROC RELIABILITY statement.

See "Analysis of Accelerated Life Test Data" on page 930 for an example using the RELATIONPLOT statement.

To create a life-stress relation plot, you must specify one *variable*. If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values. You must specify one *group-variable* to represent the values of stress. The *group-variable* must be a numeric variable.

The RELATIONPLOT statement plots the uncensored values of your data given by *variable* versus the values of the *group-variable*. You can optionally display a box-plot of the values of the data. You can also plot percentiles of the distribution fitted to the data. The RELATIONPLOT statement produces the same tabular output as the PROBPLOT statement, and all the analysis options are the same as for the PROBPLOT statement.

The elements of the RELATIONPLOT statement are described as follows.

variable
: represents the data for which a plot is to be produced. The *variable* must be a numeric variable in the input data set.

censor-variable(values)
: indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you omit the specification of *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(*variable1 variable2)*
: is another method of specifying the data for which a life-stress plot is to be produced. You can use this syntax in a situation where uncensored, interval-censored, left-censored and right-censored values occur in the same set of data. Table 30.20 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values. See "Lognormal Analysis with Arbitrary

Censoring” on page 939 for an example of using this syntax to create a probability plot.

group-variable
is a group variable. The *group-variable* must be a numeric variable in the input data set.

options
control the features of the relation plot. All *options* are specified after the slash (/) in the RELATIONPLOT statement. The “Summary of Options” section, which follows, lists all options by function.

The only type of relation plot currently available for interval data is the type in which percentiles of the fitted distribution are plotted at each stress level.

Summary of Options

Table 30.30. Analysis Options

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>CONFIDENCE=number</td><td>specifies the confidence coefficient for all confidence intervals. The number must be between 0 and 1. The default value is 0.95</td></tr>
<tr><td>CONVERGE=number</td><td>specifies the convergence criterion for maximum likelihood fit. See “Maximum Likelihood Estimation” on page 1004 for details.</td></tr>
<tr><td>CONVH=number</td><td>specifies the convergence criterion for the relative Hessian convergence criterion See “Maximum Likelihood Estimation” on page 1004 for details.</td></tr>
<tr><td>CORRB</td><td>requests parameter correlation matrix</td></tr>
<tr><td>COVB</td><td>requests parameter covariance matrix</td></tr>
<tr><td>FITTYPE=</td><td>specifies method of estimating distribution parameters</td></tr>
<tr><td>LSYX</td><td>-least squares fit to the probability plot. The probability axis is the dependent variable.</td></tr>
<tr><td>| LSXY</td><td>-least squares fit to the probability plot. The lifetime axis is the dependent variable.</td></tr>
<tr><td>| MLE</td><td>-maximum likelihood (default)</td></tr>
<tr><td>| MODEL</td><td>-use the fit from the preceding MODEL statement</td></tr>
<tr><td>| NONE</td><td>-no fit is computed</td></tr>
<tr><td>| WEIBAYES</td><td>-Weibayes method</td></tr>
<tr><td><(CONFIDENCE|CONF=<i>number</i>)></td><td>The <i>number</i> is the confidence coefficient for the Weibayes fit. The <i>number</i> is between 0 and 1, with a default value of 0.95.</td></tr>
<tr><td>ITPRINT</td><td>requests iteration history for maximum likelihood fit</td></tr>
<tr><td>LRCL</td><td>requests likelihood ratio confidence intervals for distribution parameters</td></tr>
</table>

Table 30.30. Analysis Options (continued)

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>LRCLPER</td><td>requests likelihood ratio confidence intervals for distribution percentiles</td></tr>
<tr><td>LOCATION=number < LINIT ></td><td>specifies fixed or initial value of location parameter</td></tr>
<tr><td>MAXIT=number</td><td>specifies maximum number of iterations allowed for maximum likelihood fit</td></tr>
<tr><td>NOPCTILES</td><td>suppress computation of percentiles</td></tr>
<tr><td>PCTLIST=number-list</td><td>specifies list of percentages for which to compute percentile estimates. The number-list must be a list of numbers separated by blanks or commas. Each number in the list must be between 0 and 100.</td></tr>
<tr><td>PPOS=</td><td>specifies plotting position type. See “Probability Plotting” beginning on page 995 for details.</td></tr>
<tr><td>EXPRANK
|MEDRANK
|MEDRANK1
|KM
|MKM</td><td>-expected ranks
-median ranks
-median ranks (exact formula)
-Kaplan-Meier
-modified Kaplan-Meier (default)</td></tr>
<tr><td>PPOUT</td><td>request table of cumulative probabilities</td></tr>
<tr><td>PSTABLE=number</td><td>specifies stable parameterization. The number must be between zero and one. See "Stable Parameters" on page 1007 for further information.</td></tr>
<tr><td>RELATION=ARRHENIUS |
ARRHENIUS2 |
LINEAR | POWER</td><td>specifies type of relationship between life and stress. This determines the horizontal scale used in the relation plot. See Table 30.45 on page 1007 for definitions of the transformations.</td></tr>
<tr><td>READOUT</td><td>analyzes data with readout structure</td></tr>
<tr><td>SCALE=number < SCINIT ></td><td>specifies fixed or initial value of scale parameter</td></tr>
<tr><td>SHAPE=number < SHINIT ></td><td>specifies fixed or initial value of shape parameter</td></tr>
<tr><td>SINGULAR=number</td><td>specifies singularity criterion for matrix inversion</td></tr>
<tr><td>SURVTIME=number-list</td><td>requests that survival function be computed for values in number-list</td></tr>
<tr><td>THRESHOLD=number</td><td>specifies a fixed threshold parameter. See Table 30.37 for the distributions with a threshold parameter.</td></tr>
</table>

Table 30.31. Plot Layout Options

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>CENSYMBOL=symbol | (symbol list)</td><td>specifies symbols for censored values. The symbol is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z). If you are creating overlaid plots for groups of data, you can specify different symbols for the groups with a list of symbols or letters, separated by blanks, enclosed in parentheses. If no CENSYMBOL option is specified, the symbol used for censored values is the same as for failures.</td></tr>
<tr><td>HOFFSET=value</td><td>specifies an offset for horizontal axis</td></tr>
<tr><td>INBORDER</td><td>requests a border around plots</td></tr>
<tr><td>JITTER=number</td><td>specifies amount to jitter overlaying plot symbols, in units of symbol width</td></tr>
<tr><td>LFIT=linetype</td><td>specifies a line style for fit line and confidence curves</td></tr>
<tr><td>LBOXES=number</td><td>specifies a line style for boxplots</td></tr>
<tr><td>NOCENPLOT</td><td>suppresses plotting of censored data points</td></tr>
<tr><td>NOCONF</td><td>suppresses plotting of percentile confidence curves</td></tr>
<tr><td>NOFIT</td><td>suppresses plotting of fit line and percentile confidence curves</td></tr>
<tr><td>NOFRAME</td><td>suppresses frame around plotting area</td></tr>
<tr><td>NOPPLEGEND</td><td>suppresses legend for overlaid probability plots</td></tr>
<tr><td>NORPLEGEND</td><td>suppresses legend for relation plot</td></tr>
<tr><td>PLOTDATA <DATA | MEDIANS | BOXES></td><td>requests that the data be plotted on the relationplot and specifies the representation of the data populations to be plotted</td></tr>
<tr><td>PLOTFIT <<i>number-list</i>></td><td>specifies that percentiles of the fitted distribution be plotted on the relation plot. The optional <i>number-list</i> is a list of percentiles (between 0 and 100), and, if not specified, the 50th percentile (median) is plotted.</td></tr>
<tr><td>PPLEGEND = legend-statement-name | NONE</td><td>identifies a LEGENDn statement to specify legend for overlaid probability plots</td></tr>
<tr><td>PPLOT</td><td>places a probability plot on the same page as the relation plot</td></tr>
<tr><td>RPLEGEND = legend-statement-name | NONE</td><td>identifies a LEGENDn statement to specify legend for the relation plot</td></tr>
<tr><td>TURNVLABELS</td><td>vertically strings out characters in labels for vertical axis</td></tr>
</table>

Table 30.31. Plot Layout Options (continued)

Option	**Option Description**
VOFFSET=*value*	specifies length of offset at upper end of vertical axis
WFIT=*linetype*	specifies line width for fit line and confidence curves

Table 30.32. Reference Line Options

<table>
<tr><th>Option</th><th>Option Description</th></tr>
<tr><td>HREF < (INTERSECT) >=value-list</td><td>requests reference lines perpendicular to horizontal axis. If (INTERSECT) is specified, a second reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the intersecting vertical axis reference line is labeled with the vertical axis value.</td></tr>
<tr><td>HREFLABELS=('label1' . . . 'labeln')</td><td>specifies labels for HREF= lines</td></tr>
<tr><td>HREFLABPOS=n</td><td>specifies vertical position of labels for HREF= lines. The valid values for n and the corresponding label placements are shown below.
<table>
<tr><th>n</th><th>label placement</th></tr>
<tr><td>1</td><td>top</td></tr>
<tr><td>2</td><td>staggered from top</td></tr>
<tr><td>3</td><td>bottom</td></tr>
<tr><td>4</td><td>staggered from bottom</td></tr>
<tr><td>5</td><td>alternating from top</td></tr>
<tr><td>6</td><td>alternating from bottom</td></tr>
</table></td></tr>
<tr><td>LHREF=linetype</td><td>specifies a line style for HREF= lines</td></tr>
<tr><td>LSREF=linetype</td><td>specifies a line style for SREF= lines</td></tr>
<tr><td>LVREF=linetype</td><td>specifies a line style for VREF= lines</td></tr>
<tr><td>SREF=value-list</td><td>specifies reference lines perpendicular to horizontal stress axis</td></tr>
<tr><td>SREFLABELS=('label1' . . . 'labeln')</td><td>specifies labels for SREF= lines</td></tr>
<tr><td>SREFLABPOS=n</td><td>specifies horizontal position of labels for SREF= lines. The valid values for n and the corresponding label placements are shown below.
<table>
<tr><th>n</th><th>label placement</th></tr>
<tr><td>1</td><td>top</td></tr>
<tr><td>2</td><td>staggered from top</td></tr>
<tr><td>3</td><td>bottom</td></tr>
<tr><td>4</td><td>staggered from bottom</td></tr>
</table></td></tr>
</table>

Table 30.32. Reference Line Options (continued)

Option	**Option Description**
VREF < (INTERSECT) >=*value-list*	requests reference lines perpendicular to vertical axis.If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value.
VREFLABELS=*('label1' . . . 'labeln')*	specifies labels for VREF= lines
VREFLABPOS=*n*	specifies horizontal position of labels for VREF= lines. The valid values for *n* and the corresponding label placements are shown below. *n* | **label placement** 1 | left 2 | right

Table 30.33. Text Enhancement Options

Option	**Option Description**
FONT=*font*	specifies a software font for text
HEIGHT=*value*	specifies height of text used outside framed areas
INFONT=*font*	specifies a software font for text inside framed areas
INHEIGHT=*value*	specifies height of text inside framed areas

Table 30.34. Axis Options

Option	**Option Description**

LAXIS=*value1 to value2<by value3>*	specifies tick mark values for the lifetime axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of LAXIS= lists are
	`laxis = -1 to 10`
	`laxis = 0 to 200 by 10`

Table 30.34. Axis Options (continued)

Option	Option Description
LGRID=*number*	specifies a line style for all grid lines. The *number* is between 1 and 46 and specifies a linestyle for grids.
LIFELOWER | LLOWER=*number*	specifies the lower limit on the lifetime axis scale. The LLOWER option specifies *number* as the lower lifetime axis tick mark. The tick interval and the upper lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
LIFEUPPER | LUPPER=*number*	specifies the upper limit on the lifetime axis scale. The LUPPER option specifies *number* as the upper lifetime axis tick mark. The tick interval and the lower lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
MPGRID	adds a minor grid for the probability axis
MINORLOGGRID	adds a minor grid for log axes
NOGRID	suppresses grid lines
NOLLABEL	suppresses label for life, or analysis variable, axis
NOLTICK	suppresses tick marks and tick mark labels for lifetime or analysis variable axis
NOPLABEL	suppresses label for probability axis
NOPTICK	suppresses tick marks and tick mark labels for probability axis
NOSLABEL	suppresses label for stress axis
NOSTICK	suppresses tick marks and tick mark labels for stress axis
NSTRESSTICK=*number*	specifies the number of tick intervals for stress axis for relation plot
NTICK=*number*	specifies the number of tick intervals for the lifetime axis. This option has no effect if the LAXIS option is used.
PCTLOWER | PLOWER=*number*	specifies lower limit on probability axis scale
PCTUPPER | PUPPER=*number*	specifies upper limit on probability axis scale
STRESSLOWER | SLOWER=*number*	specifies lower limit on stress axis scale
STRESSUPPER | SUPPER=*number*	specifies upper limit on stress axis scale
PAXISLABEL='*string*'	specifies label for probability axis
WAXIS=n	specifies line thickness for axes and frame

Table 30.35. Graphics Catalog Options

Option	**Option Description**
DESCRIPTION=‘*string*’	description for graphics catalog member
NAME=‘*string*’	name for plot in graphics catalog

Table 30.36. Color and Pattern Options

Option	**Option Description**
CAXIS=*color*	color for axis
CBOXES=*color*	color for box frame for boxplots
CBOXFILL=*color*	color for filling boxes for boxplots
CCENSOR=*color*	color for filling censor plot area
CENCOLOR=*color*	color for censor symbol
CFIT=*color*	color for fit line and confidence curves
CFRAME=*color*	color for frame
CGRID=*color*	color for grid lines
CHREF=*color*	color for HREF= lines
CPLOTFIT=*color*	color for percentile lines
CSREF=*color*	color for SREF= lines
CTEXT=*color*	color for text
CVREF=*color*	color for VREF= lines

UNITID Statement

UNITID *variable***;**

The UNITID statement names a *variable* in the input data set that is used to identify each individual unit in an MCFPLOT statement. The value of the UNITID variable for an observation corresponds to the name of the unit in the study for which a repair or end of history has occurred. See “Analysis of Recurrence Data on Repairs” on page 947 for an example using the UNITID statement with the MCFPLOT statement.

Details

Abbreviations and Notation

The following abbreviations and notation are used in this section.

CDF	cumulative distribution function: $F(x) = Pr\{X \leq x\}$
log	base e logarithm
$\log_{10}$	base 10 logarithm
Reliability or Survivor function	$R(x) = Pr\{X > x\}$
x_p	$p \times 100\%$ percentile: $Pr\{X \leq x_p\} = p$

Types of Lifetime Data

This section describes various types of data that you can analyze with the RELIABILITY procedure.

Lifetime data for which the values of all sample units are observed are called *complete* data. This means that the failure times are observed for all units.

Many practical problems in life data analysis involve data for which some units are unfailed. The failure time for an unfailed unit is known only to be greater than the last running time. This type of data is said to be *right censored*, and the censoring time is used in the analysis of the data. Data for which censoring times are intermixed with failure times are sometimes called *multiply censored* or *progressively censored*.

Failure times may be known only to be less than some value. This type of data is called *left censored*.

Another common situation is where the failure times of units are not known exactly, but time intervals that contain the failure times are known. This type of data is called *interval censored*.

Interval-censored data for which all units share common interval endpoints are called *readout*, *inspection*, or *grouped* data.

Probability Distributions

This section describes the probability distributions available in the RELIABILITY procedure for probability plotting and parameter estimation.

PROBPLOT and RELATIONPLOT Statements

Probability plots can be constructed for each of the probability distributions in Table 30.37. Estimates of two distribution parameters (*location* and *scale* or *scale* and *shape*) are computed by maximum likelihood or by least squares fitted to points on the probability plot. If one of the parameters is specified as fixed, the other is estimated. In addition, you can specify a fixed *threshold*, or *shift*, parameter for those distributions for which a threshold parameter is indicated in Table 30.37. If you do not specify a threshold parameter, the threshold is set to 0.

Note that you should not interpret the parameters μ and σ as representing the means and standard deviations for all of the distributions in Table 30.37. The normal is the only distribution in Table 30.37 for which this is the case.

Table 30.37. Distributions and Parameters for PROBPLOT and RELATIONPLOT Statements

Distribution	Density Function	Parameters			
		Location	Scale	Shape	Threshold
Normal	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ		
Lognormal	$\frac{1}{\sqrt{2\pi}\sigma(x-\theta)}\exp\left(-\frac{(\log(x-\theta)-\mu)^2}{2\sigma^2}\right)$	μ	σ		θ
Lognormal (base 10)	$\frac{\log(10)}{\sqrt{2\pi}\sigma(x-\theta)}\exp\left(-\frac{(\log_{10}(x-\theta)-\mu)^2}{2\sigma^2}\right)$	μ	σ		θ
Extreme Value	$\frac{1}{\sigma}\exp\left(\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)$	μ	σ		
Weibull	$\frac{\beta}{\alpha^\beta}(x-\theta)^{\beta-1}\exp\left(-\left(\frac{x-\theta}{\alpha}\right)^\beta\right)$		α	β	θ
Exponential	$\frac{1}{\alpha}\exp\left(-\left(\frac{x-\theta}{\alpha}\right)\right)$		α		θ
Logistic	$\frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left[1+\exp\left(\frac{x-\mu}{\sigma}\right)\right]^2}$	μ	σ		
Log-logistic	$\frac{\exp\left(\frac{\log(x-\theta)-\mu}{\sigma}\right)}{(x-\theta)\sigma\left[1+\exp\left(\frac{\log(x-\theta)-\mu}{\sigma}\right)\right]^2}$	μ	σ		θ

The exponential distribution shown in Table 30.37 is a special case of the Weibull distribution with $\beta = 1$. The remaining distributions in Table 30.37 are related to one another as shown in Table 30.38. The threshold parameter, θ, is assumed to be 0 in Table 30.38.

Table 30.38. Relationship among Life Distributions

Distribution of T	Parameters		Distribution of Y=logT	Parameters	
Lognormal	μ	σ	Normal	μ	σ
Weibull	α	β	Extreme Value	$\mu = \log\alpha$	$\sigma = \frac{1}{\beta}$
Log-logistic	μ	σ	Logistic	μ	σ

MODEL Statement

All of the distributions in Table 30.37 are available for regression model estimation using the MODEL statement. In addition, the generalized gamma distribution with the following probability density function $f(t)$ is available for regression model estimation in the MODEL statement.

$$f(t) = \frac{|\lambda|}{t\sigma\Gamma(\lambda^{-2})}(\lambda^{-2})^{(\lambda^{-2})}\exp\left[\lambda^{-2}\left(\lambda\left(\frac{\log(t)-\mu}{\sigma}\right)-\exp\left(\lambda\left(\frac{\log(t)-\mu}{\sigma}\right)\right)\right)\right]$$

If a lifetime T has the generalized gamma distribution, then the logarithm of the lifetime $X = \log(T)$ has the generalized log-gamma distribution, with the following probability density function $g(x)$. When the gamma distribution is specified, the logarithms of the lifetimes are used as responses, and the generalized log-gamma distribution is used to estimate the parameters by maximum likelihood.

$$g(x) = \frac{|\lambda|}{\sigma\Gamma(\lambda^{-2})}(\lambda^{-2})^{(\lambda^{-2})}\exp\left[\lambda^{-2}\left(\lambda\left(\frac{x-\mu}{\sigma}\right)-\exp\left(\lambda\left(\frac{x-\mu}{\sigma}\right)\right)\right)\right]$$

Refer to Lawless (1982, p. 240) for a description of the generalized gamma and generalized log-gamma distributions.

When $\lambda = 1$, the generalized log-gamma distribution reduces to the extreme value distribution with parameters μ and σ. In this case, the log lifetimes have the extreme value distribution, or, equivalently, the lifetimes have the Weibull distribution with parameters $\alpha = \exp(\mu)$ and $\beta = 1/\sigma$. When $\lambda = 0$, the generalized log-gamma reduces to the normal distribution with parameters μ and σ. In this case, the (unlogged) lifetimes have the lognormal distribution with parameters μ and σ. This chapter uses the notation μ for the *location*, σ for the *scale*, and λ for the *shape* parameters for the generalized log-gamma distribution.

ANALYZE Statement

You can use the ANALYZE statement to compute parameter estimates and other statistics for the distributions in Table 30.37. In addition, you can compute estimates for the binomial and Poisson distributions. The forms of these distributions are shown in Table 30.39.

Table 30.39. Binomial and Poisson Distributions

Distribution	Pr{Y=y}	Parameter	Parameter Name
Binomial	$\binom{n}{y} p^y(1-p)^{n-y}$	p	binomial probability
Poisson	$\frac{\mu^y}{y!}\exp(-\mu)$	μ	Poisson mean

Probability Plotting

Probability plots are useful tools for the display and analysis of lifetime data. Refer to Abernethy (1996) for examples using probability plots in the analysis of reliability data. Probability plots use a special scale so that a cumulative distribution function (CDF) plots as a straight line. Thus, if lifetime data are a sample from a distribution, the CDF estimated from the data plots approximately as a straight line on a probability plot for the distribution.

You can use the RELIABILITY procedure to construct probability plots for data that are complete, right censored, or interval censored (in readout form) for each of the probability distributions in Table 30.37.

A random variable Y belongs to a *location-scale* family of distributions if its CDF F is of the form

$$Pr\{Y \leq y\} = F(y) = G\left(\frac{y-\mu}{\sigma}\right)$$

where μ is the location parameter, and σ is the scale parameter. Here, G is a CDF that cannot depend on any unknown parameters, and G is the CDF of Y if $\mu = 0$ and $\sigma = 1$. For example, if Y is a normal random variable with mean μ and standard deviation σ,

$$G(u) = \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})\, du$$

and

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

Of the distributions in Table 30.37, the normal, extreme value, and logistic distributions are location-scale models. As shown in Table 30.38, if T has a lognormal, Weibull, or log-logistic distribution, then $\log(T)$ has a distribution that is a location-scale model. Probability plots are constructed for lognormal, Weibull, and log-logistic distributions by using $\log(T)$ instead of T in the plots.

Let $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ be ordered observations of a random sample with distribution function $F(y)$. A probability plot is a plot of the points $y_{(i)}$ against $m_i = G^{-1}(a_i)$, where $a_i = \hat{F}(y_i)$ is an estimate of the CDF $F(y_{(i)}) = G\left(\frac{y_{(i)}-\mu}{\sigma}\right)$. The points a_i are called *plotting positions*. The axis on which the points m_is are plotted is usually labeled with a probability scale (the scale of a_i).

If F is one of the location-scale distributions, then y is the lifetime; otherwise, the log of the lifetime is used to transform the distribution to a location-scale model.

If the data actually have the stated distribution, then $\hat{F} \approx F$,

$$m_i = G^{-1}(\hat{F}(y_i) \approx G^{-1}(G\left(\frac{y_{(i)}-\mu}{\sigma}\right)) = \frac{y_{(i)}-\mu}{\sigma}$$

and points $(y_{(i)}, m_i)$ should fall approximately on a straight line.

There are several ways to compute plotting positions from failure data. These are discussed in the next two sections.

Complete and Right-Censored Data

The censoring times must be taken into account when you compute plotting positions for right-censored data. The RELIABILITY procedure provides several methods for computing plotting positions. These are specified with the PPOS= option in the ANALYZE, PROBPLOT, and RELATIONPLOT statements. All of the methods give similar results, as illustrated in the following sections, “Expected Ranks, Kaplan-Meier, and Modified Kaplan-Meier Methods” and “Median Ranks.”

Expected Ranks, Kaplan-Meier, and Modified Kaplan-Meier Methods

Let $y_{(1)} \le y_{(2)} \le \ldots \le y_{(n)}$ be ordered observations of a random sample including failure times and censor times. Order the data in increasing order. Label all the data with reverse ranks r_i, with $r_1 = n, \ldots, r_n = 1$. For the failure corresponding to reverse rank r_i, compute the reliability

$$R_i = \left[\frac{r_i}{r_i + 1}\right] R_{i-1}$$

with $R_0 = 1$. The expected rank plotting position is computed as $a_i = 1 - R_i$. The option PPOS=EXPRANK specifies the expected rank plotting position.

For the Kaplan-Meier method,

$$R_i = \left[\frac{r_i - 1}{r_i}\right] R_{i-1}$$

The Kaplan-Meier plotting position is then computed as $a'_i = 1 - R_i$. The option PPOS=KM specifies the Kaplan-Meier plotting position.

For the modified Kaplan-Meier method, use

$$R'_i = \frac{R_i + R_{i-1}}{2}$$

where R_i is computed from the Kaplan-Meier formula with $R_0 = 1$. The plotting position is then computed as $a''_i = 1 - R'_i$. The option PPOS=MKM specifies the modified Kaplan-Meier plotting position. If the PPOS option is not specified, the modified Kaplan-Meier plotting position is used as the default method.

For complete samples, $a_i = i/(n+1)$ for the expected rank method, $a'_i = i/n$ for the Kaplan-Meier method, and $a''_i = (i - .5)/n$ for the modified Kaplan-Meier method. If the largest observation is a failure for the Kaplan-Meier estimator, then $F_n = 1$ and the point is not plotted. These three methods are shown for the field winding data in Table 30.40 and Table 30.41.

Table 30.40. Expected Rank Plotting Position Calculations

Ordered Observation	Reverse Rank	$r_i/(r_i+1)$	$\times R_{i-1}$	$= R_i$	$a_i = 1 - R_i$
31.7	16	16/17	1.0000	0.9411	0.0588
39.2	15	15/16	0.9411	0.8824	0.1176
57.5	14	14/15	0.8824	0.8235	0.1765
65.0+	13				
65.8	12	12/13	0.8235	0.7602	0.2398
70.0	11	11/12	0.7602	0.6968	0.3032
75.0+	10				
75.0+	9				
87.5+	8				
88.3+	7				
94.2+	6				
101.7+	5				
105.8	4	4/5	0.6968	0.5575	0.4425
109.2+	3				
110.0	2	2/3	0.5575	0.3716	0.6284
130.0+	1				

+ Censored Times

Table 30.41. Kaplan-Meier and Modified Kaplan-Meier Plotting Position Calculations

Ordered Observation	Reverse Rank	$(r_i-1)/r_i$	$\times R_{i-1}$	$= R_i$	$a'_i = 1 - R_i$	a''_i
31.7	16	15/16	1.0000	0.9375	0.0625	0.0313
39.2	15	14/15	0.9375	0.8750	0.1250	0.0938
57.5	14	13/14	0.8750	0.8125	0.1875	0.1563
65.0+	13					
65.8	12	11/12	0.8125	0.7448	0.2552	0.2214
70.0	11	10/11	0.7448	0.6771	0.3229	0.2891
75.0+	10					
75.0+	9					
87.5+	8					
88.3+	7					
94.2+	6					
101.7+	5					
105.8	4	3/4	0.6771	0.5078	0.4922	0.4076
109.2+	3					
110.0	2	1/2	0.5078	0.2539	0.7461	0.6192
130.0+	1					

+ Censored Times

Median Ranks

Refer to Abernethy (1996) for a discussion of the methods described in this section. Let $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ be ordered observations of a random sample including failure times and censor times. A failure order number j_i is assigned to the *i*th failure: $j_i = j_{i-1} + \Delta$, where $j_0 = 0$. The increment Δ is initially 1 and is modified when a censoring time is encountered in the ordered sample. The new increment is computed as

$$\Delta = \frac{(n+1) - \text{ previous failure order number}}{1 + \text{ number of items beyond previous censored item}}$$

The plotting position is computed for the ith failure time as

$$a_i = \frac{j_i - .3}{n + .4}$$

For complete samples, the failure order number j_i is equal to i, the order of the failure in the sample. In this case, the preceding equation for a_i is an approximation to the median plotting position computed as the median of the *i*th-order statistic from the uniform distribution on (0, 1). In the censored case, j_i is not necessarily an integer, but the preceding equation still provides an approximation to the median plotting position. The PPOS=MEDRANK option specifies the median rank plotting position.

For complete data, an alternative method of computing the median rank plotting position for failure i is to compute the exact median of the distribution of the ith order statistic of a sample of size n from the uniform distribution on (0,1). If the data are right censored, the adjusted rank j_i, as defined in the preceding paragraph, is used in place of i in the computation of the median rank. The PPOS=MEDRANK1 option specifies this type of plotting position.

Nelson (1982, p.148) provides the following example of multiply right-censored failure data for field windings in electrical generators. Table 30.42 shows the data, the intermediate calculations, and the plotting positions calculated by exact (a_i') and approximate (a_i) median ranks.

Table 30.42. Median Rank Plotting Position Calculations

Ordered Observation	Increment Δ	Failure Order Number j_i	a_i	a_i'
31.7	1.0000	1.0000	0.04268	0.04240
39.2	1.0000	2.0000	0.1037	0.1027
57.5	1.0000	3.0000	0.1646	0.1637
65.0+	1.0769			
65.8	1.0769	4.0769	0.2303	0.2294
70.0	1.0769	5.1538	0.2960	0.2953
75.0+	1.1846			
75.0+	1.3162			
87.5+	1.4808			
88.3+	1.6923			
94.2+	1.9744			
101.7+	2.3692			
105.8	2.3692	7.5231	0.4404	0.4402
109.2+	3.1590			
110.0	3.1590	10.6821	0.6331	0.6335
130.0+	6.3179			

\+ Censored Times

Interval-Censored Data

Readout Data

The RELIABILITY procedure can create probability plots for interval-censored data when all units share common interval endpoints. This type of data is called *readout* data in the RELIABILITY procedure. Estimates of the cumulative distribution function are computed at times corresponding to the interval endpoints. Right censoring can also be accommodated if the censor times correspond to interval endpoints. See “Weibull Analysis of Interval Data with Common Inspection Schedule” on page 935 for an example of a Weibull plot and analysis for interval data.

Table 30.43 illustrates the computational scheme used to compute the CDF estimates. The data are failure data for microprocessors (Nelson 1990, p.147). In Table 30.43, t_i are the interval upper endpoints, in hours, f_i is the number of units failing in interval i, and n_i is the number of unfailed units at the beginning of interval i.

Note that there is right censoring as well as interval censoring in these data. For example, two units fail in the interval (24, 48) hours, and there are 1414 unfailed units at the beginning of the interval, 24 hours. At the beginning of the next interval, (48, 168) hours, there are 573 unfailed units. The number of unfailed units that are removed from the test at 48 hours is $1414 - 2 - 573 = 839$ units. These are right-censored units.

The reliability at the end of interval i is computed recursively as

$$R_i = (1 - (f_i/n_i))R_{i-1}$$

with $R_0 = 1$. The plotting position is $a_i = 1 - R_i$.

Table 30.43. Interval-Censored Plotting Position Calculations

Interval i	Interval Endpoint t_i	f_i/n_i	$R_i' = 1-(f_i/ni)$	$R_i = R_i' R_{i-1}$	$a_i = 1 - R_i$
1	6	6/1423	0.99578	0.99578	.00421
2	12	2/1417	0.99859	0.99438	.00562
3	24	0/1415	1.00000	0.99438	.00562
4	48	2/1414	0.99859	0.99297	.00703
5	168	1/573	0.99825	0.99124	.00876
6	500	1/422	0.99763	0.98889	.01111
7	1000	2/272	0.99265	0.98162	.01838
8	2000	1/123	0.99187	0.97364	.02636

The variance $v(a_i)$ of the cumulative probability estimate $a_i = 1 - R_i$ is computed using the *exact variance* method of Nelson (1990 pp. 150-151).

If no right censoring has occurred before t_i, then a_i is a binomial probability, and exact binomial confidence limits for a_i are computed. See Binomial Distribution on page 1013 for a description of this method.

If right censoring has occurred before t_i, then two-sided approximate $100\gamma\%$ confidence limits for a_i are computed as

$$
\begin{aligned}
a_L &= a_i - K_\gamma \sqrt{v(a_i)} \\
a_U &= a_i + K_\gamma \sqrt{v(a_i)}
\end{aligned}
$$

where K_γ is the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution. The estimates a_i, confidence limits a_L and a_U, and standard errors $\sqrt{v(a_i)}$ are tabulated in the ANALYZE, PROBPLOT, and RELATIONPLOT statements for readout data. The PCONFPLT option requests that the confidence limits be displayed on probability plots.

Arbitrarily-Censored Data

The RELIABILITY procedure can create probability plots for data that consists of combinations of exact, left-censored, right-censored, and interval-censored lifetimes. Unlike the method in the previous section, failure intervals need not share common endpoints. The RELIABILITY procedure uses an iterative algorithm developed by Turnbull (1976) to compute a nonparametric maximum likelihood estimate of the cumulative distribution function for the data. Since the technique is maximum likelihood, standard errors of the cumulative probability estimates are computed from the inverse of the associated Fisher information matrix. A technique developed by Gentleman and Geyer (1994) is used to check for convergence to the maximum likelihood estimate. Also see Meeker and Escobar (1998, chap. 3) for more information.

Although this method applies to more general situations, where the intervals may be overlapping, the example of the previous section will be used to illustrate the method. Table 30.44 contains the microprocessor data of the previous section, arranged in intervals. A missing (.) lower endpoint indicates left censoring, and a missing upper endpoint indicates right censoring. These can be thought of as semi-infinite intervals with lower (upper) endpoint of negative (positive) infinity for left (right) censoring.

Table 30.44. Interval-Censored Data

Lower Endpoint	Upper Endpoint	Number Failed
.	6	6
6	12	2
24	48	2
24	.	1
48	168	1
48	.	839
168	500	1
168	.	150
500	1000	2
500	.	149
1000	2000	1
1000	.	147
2000	.	122

The following SAS program will compute the Turnbull estimate and create a lognormal probability plot.

```
data micro;
   input t1 t2 f ;
   datalines;
   . 6 6
   6 12 2
   12 24 0
   24 48 2
   24 .  1
   48 168 1
   48 .   839
   168 500 1
   168 .   150
   500 1000 2
   500 .    149
   1000 2000 1
   1000 . 147
   2000 . 122
   ;

proc reliability data=micro;
   distribution lognormal;
   freq f;
   pplot ( t1 t2 ) / itprintem
                     printprobs
                     maxitem = (1000,25)
                     nofit
                     pconfplt
                     ppout;
   run;
```

The nonparametric maximum likelihood estimate of the CDF can only increase on certain intervals, and must remain constant between the intervals. The Turnbull algorithm first computes the intervals on which the nonparametric maximum likelihood estimate of the CDF can increase. The algorithm then iteratively estimates the probability associated with each interval. The ITPRINTEM option along with the PRINTPROBS option instructs the procedure to print the intervals on which probability increases can occur and the iterative history of the estimates of the interval probabilities. The PPOUT option requests tabular output of the estimated CDF, standard errors, and confidence limits for each cumulative probability.

Figure 30.25 shows every 25th iteration and the last iteration for the Turnbull estimate of the CDF for the microprocessor data. The initial estimate assigns equal probabilities to each interval. You can specify different initial values with the PROBLIST= option. The algorithm converges in 130 iterations for this data. Convergence is determined if the change in the log-likelihood between two successive iterations less than delta, where the default value of delta is 10^{-8}. You can specify a different value for delta with the TOLLIKE= option. This algorithm is an example of an expectation-maximization (EM) algorithm. EM algorithms are known to converge slowly, but the computations within each iteration for the Turnbull algorithm are moderate. Iterations will be terminated if the algorithm does not converge after a fixed number of iterations. The default maximum number of iterations is 1000. Some data may require more iterations for convergence. You can spoecify the maximum allowed number of iterations with the MAXITEM= option in the PROBPLOT, ANALYZE, or RPLOT statements.

```
                              The RELIABILITY Procedure

            Iteration History for the Turnbull Estimate of the CDF
Iteration   Loglikelihood       (., 6)      (6, 12)     (24, 48)    (48, 168)
                            (168, 500)  (500, 1000)  (1000, 2000)    (2000, .)

        0      -1133.4051        0.125        0.125        0.125        0.125
                                 0.125         0.125         0.125        0.125
       25      -104.16622   0.00421644   0.00140548   0.00140648   0.00173338
                            0.00237846    0.00846094    0.04565407   0.93474475
       50      -101.15151   0.00421644   0.00140548   0.00140648   0.00173293
                            0.00234891    0.00727679    0.01174486   0.96986811
       75      -101.06641   0.00421644   0.00140548   0.00140648   0.00173293
                            0.00234891    0.00727127    0.00835638    0.9732621
      100      -101.06534   0.00421644   0.00140548   0.00140648   0.00173293
                            0.00234891    0.00727125    0.00801814   0.97360037
      125      -101.06533   0.00421644   0.00140548   0.00140648   0.00173293
                            0.00234891    0.00727125    0.00798438   0.97363413
      130      -101.06533   0.00421644   0.00140548   0.00140648   0.00173293
                            0.00234891    0.00727125      0.007983   0.97363551
```

Figure 30.25. Iteration History for Turnbull Estimate

If an interval probability is smaller than a tolerance (10^{-6} by default) after convergence, the probability is set to zero, the interval probabilities are renormalized so that they add to one, and iterations are restarted. Usually the algorithm converges in just a few more iterations. You can change the default value of the tolerance with the TOLPROB= option. You can specify the NOPOLISH option to avoid setting small probabilities to zero and restarting the algorithm.

If you specify the ITPRINTEM option, the table in Figure 30.26 summarizing the Turnbull estimate of the interval probabilities is printed. The columns labeled 'Reduced Gradient' and 'Lagrange Multiplier' are used in checking final convergence to the maximum likelihood estimate. The Lagrange multipliers must all be greater than or equal to zero, or the solution is not maximum likelihood. Refer to Gentleman and Geyer (1994) for more details of the convergence checking.

Lower Lifetime	Upper Lifetime	Probability	Reduced Gradient	Lagrange Multiplier
.	6	0.0042	0	0
6	12	0.0014	0	0
24	48	0.0014	0	0
48	168	0.0017	0	0
168	500	0.0023	0	0
500	1000	0.0073	-7.219342E-9	0
1000	2000	0.0080	-0.037063236	0
2000	.	0.9736	0.0003038877	0

Figure 30.26. Final Probability Estimates for Turnbull Algorithm

Figure 30.27 shows the final estimate of the CDF, along with standard errors and confidence limits. Figure 30.28 shows the CDF and pointwise confidence limits plotted on a lognormal probability plot.

Cumulative Probability Estimates

Lower Lifetime	Upper Lifetime	Cumulative Probability	95% Confidence Limits Lower	95% Confidence Limits Upper	Standard Error
6	6	0.0042	0.0019	0.0094	0.0017
12	24	0.0056	0.0028	0.0112	0.0020
48	48	0.0070	0.0038	0.0130	0.0022
168	168	0.0088	0.0047	0.0164	0.0028
500	500	0.0111	0.0058	0.0211	0.0037
1000	1000	0.0184	0.0094	0.0357	0.0063
2000	2000	0.0264	0.0124	0.0553	0.0101

Figure 30.27. Final CDF Estimates for Turnbull Algorithm

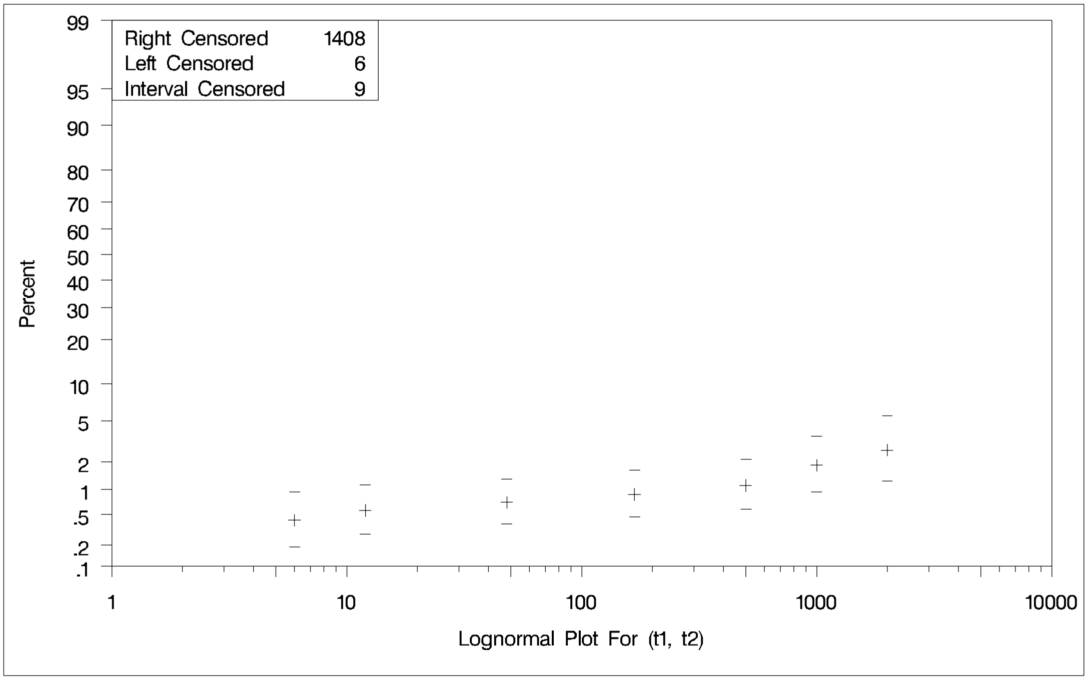

Figure 30.28. Lognormal Probability Plot for the Microprocessor Data

Parameter Estimation

Maximum Likelihood Estimation

Maximum likelihood estimation of the parameters of a statistical model involves maximizing the likelihood or, equivalently, the log likelihood with respect to the parameters. The parameter values at which the maximum occurs are the maximum likelihood estimates of the model parameters. The likelihood is a function of the parameters and of the data.

Let $x_1, x_2, \ldots, x_n$ be the observations in a random sample, including the failures and censoring times (if the data are censored). Let $f(\boldsymbol{\theta}; x)$ be the probability density of failure time, $S(\boldsymbol{\theta}; x) = Pr\{X \geq x\}$ be the reliability function, and $F(\boldsymbol{\theta}; x) = Pr\{X \leq x\}$ be the cumulative distribution function, where $\boldsymbol{\theta}$ is the vector of parameters to be estimated, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. The probability density, reliability function, and CDF are determined by the specific distribution selected as a model for the data. The log likelihood is defined as

$$
\begin{aligned}
L(\boldsymbol{\theta}) \quad = \quad & \sum_i \log(f(\boldsymbol{\theta}; x_i)) + \sum_i{}' \log(S(\boldsymbol{\theta}; x_i)) + \\
& \sum_i{}'' \log(F(\boldsymbol{\theta}; x_i)) + \sum_i{}''' [\log(F(\boldsymbol{\theta}; x_{ui}) - F(\boldsymbol{\theta}; x_{li}))]
\end{aligned}
$$

where

- $\sum$ is the sum over failed units
- $\sum'$ is the sum over right-censored units
- $\sum''$ is the sum over left-censored units
- $\sum'''$ is the sum over interval-censored units

and (x_{li}, x_{ui}) is the interval in which the ith unit is interval censored. Only the sums appropriate to the type of censoring in the data are included when the preceding equation is used.

The RELIABILITY procedure maximizes the log likelihood with respect to the parameters $\boldsymbol{\theta}$ using a Newton-Raphson algorithm. The Newton-Raphson algorithm is a recursive method for computing the maximum of a function. On the *r*th iteration, the algorithm updates the parameter vector $\boldsymbol{\theta}_r$ with

$$\boldsymbol{\theta}_{r+1} = \boldsymbol{\theta}_r - \mathbf{H}^{-1}\mathbf{g}$$

where $\mathbf{H}$ is the Hessian (second derivative) matrix, and $\mathbf{g}$ is the gradient (first derivative) vector of the log likelihood function, both evaluated at the current value of the parameter vector. That is,

$$\mathbf{g} = [g_j] = \left[\frac{\partial L}{\partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}$$

and

$$\mathbf{H} = [h_{ij}] = \left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}$$

Iteration continues until the parameter estimates converge. The convergence criterion is

$$|\theta_i^{r+1} - \theta_i^r| \le c \quad \text{if} \quad |\theta_i^{r+1}| < .01$$

$$\left|\frac{\theta_i^{r+1} - \theta_i^r}{\theta_i^{r+1}}\right| \le c \quad \text{if} \quad |\theta_i^{r+1}| \ge .01$$

for all $i = 1, 2, \ldots, p$ where c is the convergence criterion. The default value of c is 0.001, and it can be specified with the CONVERGE= option in the MODEL, PROBPLOT, RELATIONPLOT, and ANALYZE statements.

After convergence by the above criterion, the quantity

$$tc = \frac{\mathbf{g}\mathbf{H}^{-1}\mathbf{g}}{L}$$

is computed. If $tc > d$ then a warning is printed that the algorithm did not converge. tc is called the *relative Hessian* convergence criterion. The default value of d is .0001.

You can specify other values for d with the CONVH= option. The relative Hessian criterion is useful in detecting the occasional case where no progress can be made in increasing the log-likelihood, yet the gradient **g** is not zero.

A location-scale model has a CDF of the form

$$F(x) = G\left(\frac{x-\mu}{\sigma}\right)$$

where μ is the location parameter, σ is the scale parameter, and G is a standardized form $(\mu = 0, \sigma = 1)$ of the cumulative distribution function. The parameter vector is $\boldsymbol{\theta}$=(μ σ). It is more convenient computationally to maximize log likelihoods that arise from location-scale models. If you specify a distribution from Table 30.37 that is not a location-scale model, it is transformed to a location-scale model by taking the natural (base e) logarithm of the response. If you specify the lognormal base 10 distribution, the logarithm (base 10) of the response is used. The Weibull, lognormal, and log-logistic distributions in Table 30.37 are not location-scale models. Table 30.38 shows the corresponding location-scale models that result from taking the logarithm of the response.

Maximum likelihood is the default method of estimating the location and scale parameters in the MODEL, PROBPLOT, RELATIONPLOT, and ANALYZE statements. If the Weibull distribution is specified, the logarithms of the responses are used to obtain maximum likelihood estimates ($\hat{\mu}$ $\hat{\sigma}$) of the location and scale parameters of the extreme value distribution. The maximum likelihood estimates ($\hat{\alpha}$, $\hat{\beta}$) of the Weibull scale and shape parameters are computed as $\hat{\alpha} = \exp(\hat{\mu})$ and $\hat{\beta} = 1/\hat{\sigma}$.

Regression Models

In a regression model, the location parameter for the ith observation of a location-scale model is a linear function of parameters:

$$\mu_i = \mathbf{x}_i{}'\boldsymbol{\beta}$$

where $\mathbf{x}_i$ is a vector of *explanatory variables* for the ith observation determined by the experimental setup and $\boldsymbol{\beta}$ is a vector of parameters to be estimated.

You can specify a regression model using the MODEL statement. For example, if you want to relate the lifetimes of electronic parts in a test to operating temperature using the Arrhenius relationship, then an appropriate model might be

$$\mu_i = \beta_0 + x_i\beta_1$$

where $x_i = 1000/(T_i + 273.15)$, and T_i is the centigrade temperature at which the ith unit is tested. Here, $\mathbf{x}_i'$ =[1 x_i].

There are two types of explanatory variables: *continuous* variables and *class* (or *classification*) variables. Continuous variables represent physical quantities, such as temperature or voltage, and they must be numeric. Continuous explanatory variables are sometimes called *covariates*.

Class variables identify classification levels and are declared in the CLASS statement. These are also referred to as *categorical, dummy, qualitative, discrete,* or *nominal* variables. Class variables can be either character or numeric. The values of class variables are called *levels*. For example, the class variable BATCH could have levels 'batch1' and 'batch2' to identify items from two production batches. An indicator (0-1) variable is generated for each level of a class variable and is used as an explanatory variable. See Nelson (1990, p.277) for an example using an indicator variable in the analysis of accelerated life test data. In a model, an explanatory variable that is not declared in a CLASS statement is assumed to be continuous.

By default, all regression models automatically contain an intercept term; that is, the model is of the form

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \ldots$$

where β_0 does not have an explanatory variable multiplier. The intercept term can be excluded from the model by specifying the NOINT option in the MODEL statement.

For numerical stability, continuous explanatory variables are centered and scaled internally to the procedure. This transforms the parameters $\boldsymbol{\beta}$ in the original model to a new set of parameters. The parameter estimates $\boldsymbol{\beta}$ and covariances are transformed back to the original scale before reporting, so that the parameters should be interpreted in terms of the originally specified model. Covariates that are indicator variables, that is, those specified in a CLASS statement, are not centered and scaled.

Initial values of the regression parameters used in the Newton-Raphson method are computed by ordinary least squares. The parameters $\boldsymbol{\beta}$ and the scale parameter σ are jointly estimated by maximum likelihood, taking a logarithmic transformation of the responses, if necessary, to get a location-scale model.

The generalized gamma distribution is fit using the log lifetimes. The regression parameters $\boldsymbol{\beta}$, the scale parameter σ, and the shape parameter λ are jointly estimated.

The Weibull distribution shape parameter estimate is computed as $\hat{\beta} = 1/\hat{\sigma}$, where σ is the scale parameter from the corresponding extreme value distribution. The Weibull scale parameter $\hat{\alpha}_i = \exp(\mathbf{x}_i{}'\hat{\boldsymbol{\beta}})$ is not computed by the procedure. Instead, the regression parameters $\boldsymbol{\beta}$ and the shape β are reported.

In a model with a single continuous explanatory variable x, you can use the RELATION= option in the MODEL statement to specify a transformation that is applied to the variable before model fitting. Table 30.45 shows the available transformations.

Table 30.45. Variable Transformations

Relation	**Transformed variable**
ARRHENIUS (Nelson parameterization)	$1000/(x + 273.15)$
ARRHENIUS2 (activation energy parameterization)	$11605/(x + 273.15)$
POWER	$\log(x)$
LINEAR	x

Stable Parameters

The location and scale parameters (μ, σ) are estimated by maximizing the likelihood function by numerical methods, as described previously. An alternative parameterization that may have better numerical properties for heavy censoring is (η, σ), where $\eta = \mu + z_p\sigma$ and z_p is the pth quantile of the standardized distribution. See Meeker and Escobar (1998, p. 90) and Doganaksoy and Schmee (1993) for more details on alternate parameterizations.

By default, RELIABILITY estimates a value of z_p from the data that will improve the numerical properties of the estimation. You can also specify values of p from which the value of z_p will be computed with the PSTABLE= option in the ANALYZE, PROBPLOT, RELATIONPLOT, or MODEL statements. Note that a value of $p = 0.632$ for the Weibull and extreme value and $p = 0.5$ for all other distributions will give $z_p = 0$ and the parameterization will then be the usual location-scale parameterization.

All estimates and related statistics are reported in terms of the location and scale parameters (μ, σ). If you specify the ITPRINT option in the ANALYZE, PROBPLOT, or RELATIONPLOT statement, a table showing the values of p, ν, σ, and the last evaluation of the gradient and Hessian for these parameters is produced.

Covariance Matrix

An estimate of the covariance matrix of the maximum likelihood estimators (MLEs) of the parameters $\boldsymbol{\theta}$ is given by the inverse of the negative of the matrix of second derivatives of the log likelihood, evaluated at the final parameter estimates:

$$\boldsymbol{\Sigma} = [\sigma_{ij}] = -\mathbf{H}^{-1} = -\left[\frac{\partial^2 L}{\partial\theta_i \partial\theta_j}\right]^{-1}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

The negative of the matrix of second derivatives is called the Fisher information matrix. The diagonal term σ_{ii} is an estimate of the variance of $\hat{\theta}_i$. Estimates of standard errors of the MLEs are provided by

$$\mathrm{SE}_{\theta_i} = \sqrt{\sigma_{ii}}$$

An estimator of the correlation matrix is

$$\mathbf{R} = \left[\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\right]$$

The covariance matrix for the Weibull distribution parameter estimators is computed by a first-order approximation from the covariance matrix of the estimators of the corresponding extreme value parameters (μ, σ) as

$$\begin{aligned}
\mathrm{Var}(\hat{\alpha}) &= [\exp(\hat{\mu})]^2 \mathrm{Var}(\hat{\mu}) \\
\mathrm{Var}(\hat{\beta}) &= \frac{\mathrm{Var}(\hat{\sigma})}{\hat{\sigma}^4} \\
\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) &= -\frac{\exp(\hat{\mu})}{\hat{\sigma}^2}\mathrm{Cov}(\hat{\mu}, \hat{\sigma})
\end{aligned}$$

For the regression model, the variance of the Weibull shape parameter estimator $\hat{\beta}$ is computed from the variance of the estimator of the extreme value scale parameter σ

as shown previously. The covariance of the regression parameter estimator $\hat{\beta}_i$ and the Weibull shape parameter estimator $\hat{\beta}$ is computed in terms of the covariance between $\hat{\beta}_i$ and $\hat{\sigma}$ as

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}) = -\frac{\mathrm{Cov}(\hat{\beta}_i, \hat{\sigma})}{\hat{\sigma}^2}$$

Confidence Intervals for Distribution Parameters

Table 30.46 shows the method of computation of approximate two-sided $\gamma \times 100\%$ confidence limits for distribution parameters. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. In Table 30.46, K_γ represents the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution, and $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs of the location and scale parameters for the normal, extreme value, and logistic distributions. For the lognormal, Weibull, and log-logistic distributions, $\hat{\mu}$ and $\hat{\sigma}$ represent the MLEs of the corresponding location and scale parameters of the location-scale distribution that results when the logarithm of the lifetime is used as the response. For the Weibull distribution, μ and σ are the location and scale parameters of the extreme value distribution for the logarithm of the lifetime. $\mathrm{SE}_{\hat{\theta}}$ denotes the standard error of the MLE of θ, computed as the square root of the appropriate diagonal element of the inverse of the Fisher information matrix.

Table 30.46. Confidence Limit Computation

Distribution	Parameters		
	Location	Scale	Shape
Normal	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Lognormal	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Lognormal (base 10)	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Extreme Value	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Weibull		$\alpha_L = \exp[\hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})]$ $\alpha_U = \exp[\hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})]$	$\beta_L = \exp[-K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]/\hat{\sigma}$ $\beta_U = \exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]/\hat{\sigma}$
Exponential		$\alpha_L = \exp[\hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})]$ $\alpha_U = \exp[\hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})]$	
Logistic	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Log-logistic	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Generalized gamma		$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	$\mu_L = \hat{\lambda} - K_\gamma(\mathrm{SE}_{\hat{\lambda}})$ $\mu_U = \hat{\lambda} + K_\gamma(\mathrm{SE}_{\hat{\lambda}})$

Regression Parameters

Approximate $\gamma \times 100\%$ confidence limits for the regression parameter β_i are given by

$$\beta_{iL} = \hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$$

$$\beta_{iU} = \hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$$

Percentiles

The maximum likelihood estimate of the $p \times 100\%$ percentile x_p for the extreme value, normal, and logistic distributions is given by

$$\hat{x}_p = \hat{\mu} + z_p\hat{\sigma}$$

where $z_p = G^{-1}(p)$, G is the standardized CDF shown in Table 30.47, and $(\hat{\mu}, \hat{\sigma})$ are the maximum likelihood estimates of the location and scale parameters of the distribution. The maximum likelihood estimate of the percentile t_p for the Weibull, lognormal, and log-logistic distributions is given by

$$\hat{t}_p = \exp[\hat{\mu} + z_p\hat{\sigma}]$$

where $z_p = G^{-1}(p)$, and G is the standardized CDF of the location-scale model corresponding to the logarithm of the response. For the lognormal (base 10) distribution,

$$\hat{t}_p = 10^{[\hat{\mu}+z_p\hat{\sigma}]}$$

Table 30.47. Standardized Cumulative Distribution Functions

Distribution	Location-Scale Distribution	Location-Scale CDF
Weibull	Extreme Value	$1 - \exp[-\exp(z)]$
Lognormal	Normal	$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$
Log-logistic	Logistic	$\frac{\exp(z)}{1+\exp(z)}$

Confidence Intervals

The variance of the MLE of the $p \times 100\%$ percentile for the normal, extreme value, or logistic distribution is

$$Var(\hat{x}_p) = Var(\hat{\mu}) + z_p^2 Var(\hat{\sigma}) + 2Cov(\hat{\mu}, \hat{\sigma})$$

Two-sided approximate $100\gamma\%$ confidence limits for x_p are

$$\begin{aligned} x_{pL} &= \hat{x}_p - K_\gamma\sqrt{Var(\hat{x}_p)} \\ x_{pU} &= \hat{x}_p + K_\gamma\sqrt{Var(\hat{x}_p)} \end{aligned}$$

where K_γ represents the $100(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution.

The limits for the lognormal, Weibull, or log-logistic distributions are

$$\begin{aligned} t_{pL} &= \exp\left(\hat{x}_p - K_\gamma\sqrt{Var(\hat{x}_p)}\right) \\ t_{pU} &= \exp\left(\hat{x}_p + K_\gamma\sqrt{Var(\hat{x}_p)}\right) \end{aligned}$$

where x_p refers to the percentile of the corresponding location-scale distribution (normal, extreme value, or logistic) for the logarithm of the lifetime. For the lognormal (base 10) distribution,

$$t_{pL} = 10^{\left(\hat{x}_p - K_\gamma \sqrt{Var(\hat{x}_p)}\right)}$$
$$t_{pU} = 10^{\left(\hat{x}_p + K_\gamma \sqrt{Var(\hat{x}_p)}\right)}$$

Reliability Function

For the extreme value, normal, and logistic distributions shown in Table 30.47, the maximum likelihood estimate of the reliability function $R(x) = Pr\{X > x\}$ is given by

$$\hat{R}(x) = 1 - F\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right)$$

The MLE of the CDF is $\hat{F}(x) = 1 - \hat{R}(x)$.

Confidence Intervals

Let $\hat{u} = \frac{x-\hat{\mu}}{\hat{\sigma}}$. The variance of u is

$$Var(\hat{u}) \approx \frac{Var(\hat{\mu}) + \hat{u}^2 Var(\hat{\sigma}) + 2\hat{u}Cov(\hat{\mu}, \hat{\sigma})}{\hat{\sigma}^2}$$

Two-sided approximate $\gamma \times 100\%$ confidence intervals for $R(x)$ are computed as

$$R_L(x) = \hat{R}(u_2)$$

$$R_U(x) = \hat{R}(u_1)$$

where

$$u_1 = \hat{u} - K_\gamma \sqrt{Var(\hat{u})}$$

$$u_2 = \hat{u} + K_\gamma \sqrt{Var(\hat{u})}$$

and K_γ represents the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution.

The corresponding limits for the CDF are

$$F_L(x) = 1 - R_U(x)$$

$$F_U(x) = 1 - R_L(x)$$

Limits for the Weibull, lognormal, and log-logistic reliability function $R(t)$ are the same as those for the corresponding extreme value, normal, or logistic reliability $R(y)$, where $y = \log(t)$.

Estimation with the Binomial and Poisson Distributions

In addition to estimating the parameters of the distributions in Table 30.37, you can estimate parameters, compute confidence limits, compute predicted values and prediction limits, and compute chi-squared tests for differences in groups for the binomial and Poisson distributions using the ANALYZE statement. Specify either BINOMIAL or POISSON in the DISTRIBUTION statement to use one of these distributions. The ANALYZE statement options available for the binomial and Poisson distributions are given in Table 30.5. See “Analysis of Binomial Data” on page 952 for an example of an analysis of binomial data.

Binomial Distribution

If r is the number of successes and n is the number of trials in a binomial experiment, then the maximum likelihood estimator of the probability p in the binomial distribution in Table 30.39 is computed as

$$\hat{p} = r/n$$

Two-sided $\gamma \times 100\%$ confidence limits for p are computed as in Johnson, Kotz, and Kemp (1992, p.130):

$$p_L = \frac{\nu_1 F[(1-\gamma)/2; \nu_1, \nu_2]}{\nu_2 + \nu_1 F[(1-\gamma)/2; \nu_1, \nu_2]}$$

with $\nu_1 = 2r$ and $\nu_2 = 2(n - r + 1)$ and

$$p_U = \frac{\nu_1 F[(1+\gamma)/2; \nu_1, \nu_2]}{\nu_2 + \nu_1 F[(1+\gamma)/2; \nu_1, \nu_2]}$$

with $\nu_1 = 2(r+1)$ and $\nu_2 = 2(n-r)$, where $F[\gamma; \nu_1, \nu_2]$ is the $\gamma \times 100\%$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator.

You can compute a sample size required to estimate p within a specified tolerance w with probability γ. Nelson (1982, p. 206) gives the following formula for the approximate sample size:

$$n \approx \hat{p}(1-\hat{p}) \left(\frac{K_\gamma}{w} \right)^2$$

where K_γ is the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution. The formula is based on the normal approximation for the distribution of $\hat{p}$. Nelson recommends using this formula if $np > 10$ and $np(1-p) > 10$. The value of γ used for computing confidence limits is used in the sample size computation. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. You specify a tolerance of *number* with the TOLERANCE(*number*) option.

The predicted number of successes X in a future sample of size m, based on the previous estimate of p, is computed as

$$\hat{X} = m(r/n) = m\hat{p}$$

Two-sided approximate $\gamma \times 100)\%$ prediction limits are computed as in Nelson (1982, p. 208). The prediction limits are the solutions X_L and X_U of

$$X_U/m = [(r+1)/n]F[(1+\gamma)/2; 2(r+1), 2X_U]$$

$$m/(X_L+1) = (n/r)F[(1+\gamma)/2; 2(X_L+1), 2r]$$

where $F[\gamma; \nu_1, \nu_2]$ is the $\gamma \times 100\%$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator. You request predicted values and prediction limits for a future sample of size *number* with the PREDICT(*number*) option.

You can test groups of binomial data for equality of their binomial probability using the ANALYZE statement. You specify the K groups to be compared with a group variable having K levels.

Nelson (1982, p.450) discusses a chi-squared test statistic for comparing K binomial proportions for equality. Suppose there are r_i successes in n_i trials for $i = 1, 2, \ldots, K$. The grouped estimate of the binomial probability is

$$\hat{p} = \frac{r_1 + r_2 + \cdots + r_K}{n_1 + n_2 + \cdots + n_K}$$

The chi-squared test statistic for testing the hypothesis $p_1 = p_2 = \ldots = p_K$ against $p_i \neq p_j$ for some i and j is

$$Q = \sum_{i=1}^{K} \frac{(r_i - n_i\hat{p})^2}{n_i\hat{p}(1-\hat{p})}$$

The statistic Q has an asymptotic chi-squared distribution with $K-1$ degrees of freedom. The RELIABILITY procedure computes the contribution of each group to Q, the value of Q, and the *p*-value for Q based on the limiting chi-squared distribution with $K-1$ degreees of freedom. If you specify the PREDICT option, predicted values and prediction limits are computed for each group, as well as for the pooled group. The *p*-value is defined as $p_0 = 1 - \chi^2_{K-1}[Q]$, where $\chi^2_{K-1}[x]$ is the chi-squared CDF with $K-1$ degrees of freedom, and Q is the observed value. A test of the hypothesis of equal binomial probabilities among the groups with significance level α is

- $p_0 > \alpha$: do not reject the equality hypothesis
- $p_0 \leq \alpha$: reject the equality hypothesis

Poisson Distribution

You can use the ANALYZE statement to model data using the Poisson distribution. The data consists of a count Y of occurrences in a "length" of observation T. Observation T is typically an *exposure time*, but it can have other units, such as distance. The ANALYZE statement enables you to compute the rate of occurrences, confidence limits, and prediction limits.

An estimate of the rate λ is computed as

$$\hat{\lambda} = Y/T$$

Two-sided $\gamma \times 100\%$ confidence limits for λ are computed as in Nelson (1982, p. 201):

$$\lambda_L = .5\chi^2[(1-\gamma)/2; 2Y]/T$$

$$\lambda_U = .5\chi^2[(1+\gamma)/2; 2(Y+1)]/T$$

where $\chi^2[\delta;\nu]$ is the $\delta \times 100\%$ percentile of the chi-squared distribution with ν degrees of freedom.

You can compute a length T required to estimate λ within a specified tolerance w with probability γ. Nelson (1982, p. 202) provides the following approximate formula:

$$\hat{T} \approx \hat{\lambda}\left(\frac{K_\gamma}{w}\right)^2$$

where K_γ is the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution. The formula is based on the normal approximation for $\hat{\lambda}$ and is more accurate for larger values of λT. Nelson recommends using the formula when $\lambda T > 10$. The value of γ used for computing confidence limits is also used in the length computation. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. You specify a tolerance of *number* with the TOLERANCE(*number*) option.

The predicted future number of occurrences in a length S is

$$\hat{X} = (Y/T)S = \hat{\lambda}S$$

Two-sided approximate $\gamma \times 100\%$ prediction limits are computed as in Nelson (1982, p. 203). The prediction limits are the solutions X_L and X_U of

$$X_U/S = [(Y+1)/T]F[(1+\gamma)/2; 2(Y+1), 2X_U]$$

$$S/(X_L+1) = (T/Y)F[(1+\gamma)/2; 2(X_L+1), 2Y]$$

where $F[\gamma;\nu_1,\nu_2]$ is the $\gamma \times 100\%$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator. You request predicted values and prediction limits for a future exposure *number* with the PREDICT(*number*) option.

You can compute a chi-squared test statistic for comparing K Poisson rates for equality. You specify the K groups to be compared with a group variable having K levels.

Refer to Nelson (1982, p.444) for more information on this test. Suppose that there are Y_i Poisson counts in lengths T_i for $i = 1, 2, \ldots, K$ and that the Y_i are independent. The grouped estimate of the Poisson rate is

$$\hat{\lambda} = \frac{Y_1 + Y_2 + \cdots + Y_K}{T_1 + T_2 + \cdots + T_K}$$

The chi-squared test statistic for testing the hypothesis $\lambda_1 = \lambda_2 = \ldots = \lambda_K$ against $\lambda_i \neq \lambda_j$ for some i and j is

$$Q = \sum_{i=1}^{K} \frac{(Y_i - \hat{\lambda}T_i)^2}{\hat{\lambda}T_i}$$

The statistic Q has an asymptotic chi-squared distribution with $K - 1$ degrees of freedom. The RELIABILITY procedure computes the contribution of each group to Q, the value of Q, and the *p*-value for Q based on the limiting chi-squared distribution with $K-1$ degreees of freedom. If you specify the PREDICT option, predicted values and prediction limits are computed for each group, as well as for the pooled group. The *p*-value is defined as $p_0 = 1 - \chi^2_{K-1}[Q]$, where $\chi^2_{K-1}[x]$ is the chi-squared CDF with $K - 1$ degrees of freedom and Q is the observed value. A test of the hypothesis of equal Poisson rates among the groups with significance level α is

- $p_0 > \alpha$: accept the equality hypothesis
- $p_0 \leq \alpha$: reject the equality hypothesis

Least Squares Fit to the Probability Plot

Fitting to the probability plot by least squares is an alternative to maximum likelihood estimation of the parameters of a life distribution. Only the failure times are used. A least squares fit is computed using points $(x_{(i)}, m_i)$, where $m_i = F^{-1}(a_i)$ and a_i are the plotting positions as defined in
sref[d]ppos. The x_i are either the lifetimes for the normal, extreme value, or logistic distributions or the log lifetimes for the lognormal, Weibull, or log-logistic distributions. The ANALYZE, PROBPLOT, or RELATIONPLOT statement option FITTYPE=LSXY specifies the $x_{(i)}$ as the dependent variable ('y-coordinate') and the m_i as the independent variable ('x-coordinate'). You can optionally reverse the quantities used as dependent and independent variables by specifying the FITTYPE=LSYX option.

Weibayes Estimation

Weibayes estimation is a method of performing a Weibull analysis when there are few or no failures. The FITTYPE=WEIBAYES option requests this method. The method of Nelson (1985) is used to compute a one-sided confidence interval for the

Weibull scale parameter when the Weibull shape parameter is specified. Also refer to Abernethy (1996) for more discussion and examples. The Weibull shape parameter β is assumed to be known and is specified to the procedure with the SHAPE=*number* option. Let $T_1, T_2, \ldots, T_n$ be the failure and censoring times, and let $r \geq 0$ be the number of failures in the data. If there are no failures $(r = 0)$, a lower $\gamma \times 100\%$ confidence limit for the Weibull scale parameter α is computed as

$$\alpha_L = \{\sum_{i=1}^{n} T_i^{\beta} / [-\log(1-\gamma)]\}^{1/\beta}$$

The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option.

If $r \geq 1$, the MLE of α is given by

$$\hat{\alpha} = [\sum_{i=1}^{n} T_i^{\beta} / r]^{1/\beta}$$

and a lower $\gamma \times 100\%$ confidence limit for the Weibull scale parameter α is computed as

$$\alpha_L = \hat{\alpha}[2r/\chi^2(\gamma, 2r+2)]^{1/\beta}$$

where $\chi^2(\gamma, 2r+2)$ is the γ percentile of a chi-square distribution with $2r+2$ degrees of freedom. The procedure uses the specified value of β and the computed value of α_L to compute distribution percentiles and the reliability function.

Regression Model Observation-Wise Statistics

For regression models that are fit using the MODEL statement, you can specify a variety of statistics to be computed for each observation in the input data set. This section describes the method of computation for each statistic. See Table 30.21 and Table 30.22 for the syntax for requesting these statistics.

Predicted Values

The linear predictor is

$$\hat{\mu}_i = \mathbf{x}_i{}'\boldsymbol{\beta}$$

where $\mathbf{x}_i$ is the vector of explanatory variables for the ith observation.

Percentiles

An estimator of the $p \times 100\%$ percentile x_p for the ith observation for the extreme value, normal, and logistic distributions is

$$\hat{x}_{i,p} = \mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + z_p\hat{\sigma}$$

where $z_p = G^{-1}(p)$, G is the standardized CDF, and σ is the distribution scale parameter.

An estimator of the $p \times 100\%$ percentile t_p for the ith observation for the Weibull, lognormal, and log-logistic distributions is

$$\hat{t}_{i,p} = \exp[\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + z_p\hat{\sigma}]$$

where G is the standardized CDF of the extreme value, normal, or logistic distribution that corresponds to the logarithm of the lifetime, and σ is the distribution scale parameter.

The percentile of the lognormal (base 10) distribution is

$$\hat{t}_{i,p} = 10^{[\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + z_p\hat{\sigma}]}$$

where G is the CDF of the standard normal distribution.

An estimator of the $p \times 100\%$ percentile t_p for the ith observation for the generalized gamma distribution is

$$\hat{t}_{i,p} = \exp[\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + w_{\lambda,p}\hat{\sigma}]$$

where

$$w_{\lambda,p} = \frac{1}{\lambda}\log\left(\frac{\lambda^2}{2}\chi^2_{(2/\lambda^2),p}\right)$$

and $\chi^2_{k,p}$ is the $p \times 100\%$ percentile of the chi-squared distribution with k degrees of freedom.

Standard Errors of Percentile Estimator

For the extreme value, normal, and logistic distributions, the standard error of the estimator of the $p \times 100\%$ percentile is computed as

$$\sigma_{i,p} = \sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_i \\ z_p \end{bmatrix}$$

and $\boldsymbol{\Sigma}$ is the covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$.

For the Weibull, lognormal, and log-logistic distributions, the standard error is computed as

$$\sigma_{i,p} = \exp(x_{i,p})\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

where $x_{i,p}$ is the percentile computed from the extreme value, normal, or logistic distribution that corresponds to the logarithm of the lifetime. The standard error for the lognormal (base 10) distribution is computed as

$$\sigma_{i,p} = 10^{x_{i,p}}\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

The standard error for the generalized gamma distribution percentile is computed as

$$\sigma_{i,p} = \exp[\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + w_{\lambda,p}\hat{\sigma}]\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_i \\ w_{\lambda,p} \\ \hat{\sigma}\frac{\partial w_{\lambda,p}}{\partial \lambda} \end{bmatrix}$$

$\boldsymbol{\Sigma}$ is the covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\sigma}, \hat{\lambda})$, $\boldsymbol{\beta}$ is the vector of regression parameters, σ is the scale parameter, and λ is the shape parameter.

Confidence Limits for Percentiles

Two-sided approximate $100\gamma\%$ confidence limits for $x_{i,p}$ for the extreme value, normal, and logistic distributions are computed as

$$\begin{aligned} x_L &= \hat{x}_{i,p} - K_\gamma \sigma_{i,p} \\ x_U &= \hat{x}_{i,p} + K_\gamma \sigma_{i,p} \end{aligned}$$

where K_γ represents the $100(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution.

Limits for the Weibull, lognormal, and log-logistic percentiles are computed as

$$\begin{aligned} t_L &= \exp(x_L) \\ t_U &= \exp(x_U) \end{aligned}$$

where x_L and x_U are computed from the corresponding distributions for the logarithms of the lifetimes. For the lognormal (base 10) distribution,

$$\begin{aligned} t_L &= 10^{x_L} \\ t_U &= 10^{x_U} \end{aligned}$$

Limits for the generalized gamma distribution percentiles are computed as

$$t_L = \exp\left[\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + w_{\lambda,p}\hat{\sigma} - K_\gamma\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}\right]$$
$$t_U = \exp\left[\mathbf{x}_i{}'\hat{\boldsymbol{\beta}} + w_{\lambda,p}\hat{\sigma} + K_\gamma\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}\right]$$

Reliability Function

For the extreme value, normal, and logistic distributions, an estimate of the reliability function evaluated at the response y_i is computed as

$$R(y_i) = 1 - G\left(\frac{y_i - \mathbf{x}_i{}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)$$

where $G(x)$ is the standardized CDF of the distribution from Table 30.47.

Estimates of the reliability function evaluated at the response t_i for the Weibull, lognormal, log-logistic, and generalized gamma distributions are computed as

$$R(t_i) = 1 - G\left(\frac{\log(t_i) - \mathbf{x}_i{}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)$$

where $G(x)$ is the standardized CDF of the corresponding extreme value, normal, logistic, or generalized log-gamma distributions.

Residuals

The RELIABILITY procedure computes several different kinds of residuals. In the following equations, y_i represents the ith response value if the extreme value, normal, or logistic distributions are specified. If t_i is the ith response and if the Weibull, lognormal, log-logistic, or generalized gamma distributions are specified, then y_i represents the logarithm of the response $y_i = \log(t_i)$. If the lognormal (base 10) distribution is specified, then $y_i = \log_{10}(t_i)$.

Raw Residuals

The raw residual is computed as

$$r_{Ri} = y_i - \mathbf{x}_i{}'\hat{\boldsymbol{\beta}}$$

Standardized Residuals

The standardized residual is computed as

$$r_{Si} = \frac{y_i - \mathbf{x}_i{}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}$$

Adjusted Residuals

If an observation is right censored, then the standardized residual for that observation is also right censored. Adjusted residuals adjust censored standardized residuals upward by adding a percentile of the residual lifetime distribution, given that the standardized residual exceeds the censoring value. The default percentile is the median (50th percentile), but you can, optionally, specify a $\gamma \times 100\%$ percentile using the RESIDALPHA=γ option in MODEL statement. The $\gamma \times 100$ percentile residual life is computed as in Joe and Proschan (1984). The adjusted residual is computed as

$$r_{Ai} = \begin{cases} G^{-1}[1 - (1-\gamma)S(u_i)] & \text{for right-censored observations} \\ u_i & \text{for uncensored observations} \end{cases}$$

where G is the standard CDF,

$$S(u) = 1 - G(u)$$

is the reliability function, and

$$u_i = \frac{y_i - \mathbf{x}_i{}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}$$

If the generalized gamma distribution is specified, the standardized CDF and reliability functions include the estimated shape parameter $\hat{\lambda}$.

Modified Cox-Snell Residuals

Let

$$\delta_i = \begin{cases} 1 & \text{for uncensored observations} \\ 0 & \text{for right-censored observations} \end{cases}$$

The Cox-Snell residual is defined as

$$r_{Ci} = -\log(R(y_i))$$

where

$$R(y) = 1 - G\left(\frac{y - \mathbf{x}_i{}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)$$

is the reliability function. The modified Cox-Snell residual is computed as in Collett (1994, p.152):

$$r'_{Ci} = r_{Ci} + (1 - \delta_i)\alpha$$

where α is an adjustment factor. If the fitted model is correct, the Cox-Snell residual has approximately a standard exponential distribution for uncensored observations. If an observation is censored, the residual evaluated at the censoring time is not as large as the residual evaluated at the (unknown) failure time. The adjustment factor α adjusts the censored residuals upward to account for the censoring. The default is $\alpha = 0.693$, the median of the standard exponential distribution. You can, optionally, specify any adjustment factor by using the MODEL statement option RESIDADJ=α. Another commonly used value is the mean of the standard exponential distribution, $\alpha = 1$.

Deviance Residuals

Deviance residuals are a zero-mean, symmetrized version of modified Cox-Snell residuals. Deviance residuals are computed as in Collett (1994, p.153):

$$r_{Di} = \mathrm{sgn}(\delta_i - r_{Ci})\{-2[\delta_i - r_{Ci} + \delta_i \log(r_{Ci})]\}^{1/2}$$

where

$$\mathrm{sgn}(u) = \begin{cases} -1 & \text{if } u < 0 \\ 1 & \text{if } u \geq 0 \end{cases}$$

Recurrence Data from Repairable Systems

When a repairable system fails, it is repaired and placed back in service. As a repairable system ages, it accumulates a history of repairs and costs of repairs. The mean cumulative function (MCF) $M(t)$ is defined as the population mean of the cumulative number (or cost) of repairs up until time t. You can use the RELIABILITY procedure to compute and plot nonparametric estimates and plots of the MCF for the number of repairs or the cost of repairs. The Nelson (1995) confidence limits for the MCF are also computed and plotted. You can compute and plot estimates of the difference of two MCFs and confidence intervals. This is useful for comparing the repair performance of two systems. See "Analysis of Recurrence Data on Repairs" on page 947 and "Comparison of Two Samples of Repair Data" on page 949 for examples of the analysis of recurrence data from repairable systems.

Refer to Nelson (1995), Nelson (1988), Doganaksoy and Nelson (1991), and Nelson and Doganaksoy (1989) for discussions and examples of repairable systems analysis.

Formulas for the MCF estimator $\hat{M}(t)$ and the variance of the estimator $\mathrm{Var}(\hat{M}(t))$ are given in Nelson (1995). Table 30.48 shows a set of artificial repair data from Nelson (1988). For each system, the data consist of the system and cost for each repair. If you want to compute the MCF for the number of repairs, rather than cost of repairs, then you should set the cost equal to 1 for each repair. A plus sign (+) in place of a cost indicates that the age is a censoring time. The repair history of each system ends with a censoring time.

Table 30.48. System Repair Histories for Artificial Data

Unit	(Age in Months, Cost in $100)			
6	(5,$3)	(12,$1)	(12,+)	
5	(16,+)			
4	(2,$1)	(8,$1)	(16,$2)	(20,+)
3	(18,$3)	(29,+)		
2	(8,$2)	(14,$1)	(26,$1)	(33,+)
1	(19,$2)	(39,$2)	(42,+)	

Table 30.49 illustrates the calculation of the MCF estimate from the data in Table 30.48. The RELIABILITY procedure uses the following rules for computing the MCF estimates.

1. Order all events (repairs and censoring) by age from smallest to largest.
   - If the event ages of the same or different systems are equal, the corresponding data are sorted from the largest repair cost to the smallest. Censoring events always sort as smaller than repair events with equal ages.
   - When event ages and values of more than one system coincide, the corresponding data are sorted from the largest system identifier to the smallest. The system IDs can be numeric or character, but they are always sorted in ASCII order.
2. Compute the number of systems I in service at the current age as the number in service at the last repair time minus the number of censored units in the intervening times.
3. For each repair, compute the mean cost as the cost of the current repair divided by the number in service I.
4. Compute the MCF for each repair as the previous MCF plus the mean cost for the current repair.

Table 30.49. Calculation of MCF for Artificial Data

Event	(Age,Cost)	Number I in Service	Mean Cost	MCF
1	(2,\$1)	6	\$1/6=0.17	0.17
2	(5,\$3)	6	\$3/6=0.50	0.67
3	(8,\$2)	6	\$2/6=0.33	1.00
4	(8,\$1)	6	\$1/6=0.17	1.17
5	(12,\$1)	6	\$1/6=0.17	1.33
6	(12,+)	5		
7	(14,\$1)	5	\$1/5=0.20	1.53
8	(16,\$2)	5	\$2/5=0.40	1.93
9	(16,+)	4		
10	(18,\$3)	4	\$3/4=0.75	2.68
11	(19,\$2)	4	\$2/4=0.50	3.18
12	(20,+)	3		
13	(26,\$1)	3	\$1/3=0.33	3.52
14	(29,+)	2		
15	(33,+)	1		
16	(39,\$2)	1	\$2/1=2.00	5.52
17	(42,+)	0		

The variance of the estimator of the MCF $\text{Var}(\hat{M}(t))$ is computed as in Nelson (1995). If the VARMETHOD2 option is specified, the method of Lawless and Nadeau (1995) is used to compute the variance of the estimator of the MCF. This method is recommended if the number of systems or events is large. Approximate two-sided $\gamma \times 100\%$ confidence limits for $M(t)$ are computed as

$$M_L(t) = \hat{M}(t) - K_\gamma \sqrt{\text{Var}(\hat{M}(t))}$$

$$M_U(t) = \hat{M}(t) + K_\gamma \sqrt{\text{Var}(\hat{M}(t))}$$

where K_γ represents the $100(1+\gamma)/2$ percentile of the standard normal distribution.

Figure 30.29 displays the tabular output produced by the RELIABILITY procedure for the artificial data. The first table in Figure 30.29 displays the input data set, the number of observations used in the analysis, the number of systems (units), and the number of repair events. The second table displays the system age, MCF estimate, standard error, approximate confidence limits, and system ID for each event.

The RELIABILITY Procedure

Repair Data Summary

Input Data Set	WORK.MCFART
Observations Used	17
Number of Units	6
Number of Events	11

Repair Data Analysis

Age	Sample MCF	Standard Error	95% Confidence Limits Lower	Upper	Unit ID
2.00	0.167	0.167	-0.160	0.493	sys4
5.00	0.667	0.494	-0.302	1.636	sys6
8.00	1.000	0.516	-0.012	2.012	sys2
8.00	1.167	0.543	0.103	2.230	sys4
12.00	1.333	0.667	0.027	2.640	sys6
12.00	.	.	.	.	sys6
14.00	1.533	0.764	0.035	3.032	sys2
16.00	1.933	0.951	0.069	3.797	sys4
16.00	.	.	.	.	sys5
18.00	2.683	0.913	0.894	4.473	sys3
19.00	3.183	0.641	1.926	4.440	sys1
20.00	.	.	.	.	sys4
26.00	3.517	0.679	2.185	4.848	sys2
29.00	.	.	.	.	sys3
33.00	.	.	.	.	sys2
39.00	5.517	0.679	4.185	6.848	sys1
42.00	.	.	.	.	sys1

Figure 30.29. PROC RELIABILITY Output for the Artificial Data

Estimates of the difference between two MCFs $\text{MDIFF}(t) = M_1(t) - M_2(t)$ and the variance of the estimator are computed as in Doganaksoy and Nelson (1991). Confidence limits for the MCF difference function are computed in the same way as for the MCF, using the variance of the MCF difference function estimator.

ODS Table Names

The following tables contain the ODS table names created by the RELIABILITY Procedure, organized by the statements that produce them.

Table 30.50. Tables Produced with the ANALYZE Statement

Table Name	Description
ConvergenceStatus	convergence status
CorrMat	parameter correlation matrix
CovMat	parameter covariance matrix
DatSum	summary of fit
GradHess	last evaluation of parameters, gradient, and Hessian
IterEM	iteration history for Turnbull algorithm
IterLRParm	iteration history for likelihood ratio confidence intervals for parameters
IterLRPer	iteration history for likelihood ratio confidence intervals for percentiles
IterParms	iteration history for parameter estimates
Lagrange	Lagrange multiplier statistics
PBEst	Poisson/binomial estimates by group
PBPred	Poisson/binomial predicted values
PBPredTol	Poisson/binomial predicted values by group
PBSum	Poisson/binomial analysis summary
PBTol	Poisson/binomial tolerance estimates
PctEst	percentile estimates
ParmEst	parameter estimates
ParmOther	fitted distribution mean, median, mode
PGradHess	last evaluation of parameters, gradient, and Hessian in terms of stable parameters
ProbabilityEstimates	nonparametric cumulative distribution function estimates
RelInfo	model information
SurvEst	survival function estimates
TurnbullGrad	interval probabilities, reduced grdient, Lagrange multipliers for Turnbull algorithm
WCorrMat	parameter correlation matrix for Weibull distribution
WCovMat	parameter covariance matrix for Weibull distribution

Table 30.51. Tables Produced with the MCFPLOT Statement

Table Name	Description
McfDEst	MCF difference estimates
McfDSum	MCF difference data summary
McfEst	MCF estimates
McfSum	MCF data summary

Table 30.52. Tables Produced with the MODEL Statement

Table Name	Description
MConvergenceStatus	convergence status
ModClassLevels	class level information
ModCorMat	parameter correlation matrix
ModCovMat	parameter covariance matrix
ModFitSum	summary of fit
ModInfo	model information
ModIterLRparm	iteration history for likelihood ratio confidence intervals for parameters
ModIterParms	iteration history for parameter estimates
ModLagr	Lagrange multiplier statistics
ModLastGradHess	last evaluation of the gradient and Hessian
ModObstats	observation statistics
ModParmInfo	parameter information
ModPrmEst	parameter estimates

Table 30.53. Tables Produced with PROBPLOT and RELATIONPLOT Statements

Table Name	Description
ConvergenceStatus	convergence status
CorrMat	parameter correlation matrix
CovMat	parameter covariance matrix
DatSum	summary of fit
GradHess	last evaluation of parameters, gradient, and Hessian
IterEM	iteration history for Turnbull algorithm
IterLRParm	iteration history for likelihood ratio confidence intervals for parameters
IterLRPer	iteration history for likelihood ratio confidence intervals for percentiles
IterParms	iteration history for parameter estimates
Lagrange	Lagrange multiplier statistics
PctEst	percentile estimates
ParmEst	parameter estimates
ParmOther	fitted distribution mean, median, mode
PGradHess	last evaluation of parameters, gradient, and Hessian in terms of stable parameters
ProbabilityEstimates	nonparametric cumulative distribution function estimates
RelInfo	model information
SurvEst	survival function estimates
TurnbullGrad	interval probabilities, reduced grdient, Lagrange multipliers for Turnbull algorithm
WCorrMat	parameter correlation matrix for Weibull distribution
WCovMat	parameter covariance matrix for Weibull distribution

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC® User's Guide, Version 8

ISBN 1–58025–493–4

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC® User's Guide, Version 8

ISBN 1–58025–493–4

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

The Institute is a private company devoted to the support and further development of its software and related services.